COMPARATIVE ANALYSIS OF STEM CARBON PARTITIONING AND DEVELOPMENT IN ANDROPOGONEAE BIOENERGY GRASSES

BY

BRANDON T. JAMES

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Stephen Moose, Chair
Professor Matthew Hudson
Assistant Professor Yoshie Hanzawa
Gutgsell Endowed Professor Stephen Long

ABSTRACT

Plants fix carbon during photosynthetic reactions but how this carbon is distributed and stored throughout the plant greatly depends on the species. While cellulose and lignin are the more stable forms, the plant only needs to produce enough of it to allow for basic cell wall function. Starch is the most common storage form for excess sugar accumulation; however, plants can also store carbon in the form of sucrose even though it is highly unstable and makes them a target for pests. Typically sucrose is stored in the vacuoles of plants but some species also store sucrose in the storage parenchyma. Members of the Andropogoneae tribe display a good example of this versatility in carbon storage. This tribe of grasses is known for their high levels of biomass production and includes many prominent crops such as sorghum, maize, and sugarcane and the emerging biofuel feedstock miscanthus. These crops are closely related and not only handle carbon storage in different ways but also have different carbon storage forms within each species. There are varieties in each species that store sugar with sugarcane being the most recognized for its high levels of sucrose in the stem. There are also varieties in sorghum and maize that funnel their carbon into storage for grain production. Stem fiber can also be a sink for carbon as seen by the high biomass varieties of Saccharum energy cane. Miscanthus stores large volumes of carbon as above ground cellulo-lignosic biomass during the growing season and then begins funneling carbon in the form of starch into its rhizomes during the fall for overwintering. The differences in carbon partitioning among species or genotypes are likely mediated by changes in the expression of genes that control developmental programs or carbon metabolism. Maize and sorghum have sequenced genomes, but the complex polyploid genome of Saccharum presents challenges for studies of gene function and regulation. Gene expression profiling is an efficient first step to gain an understanding of the biological pathways whose activities are associated with phenotypic variation in stem carbon partitioning within Andropogoneae grasses. This study conducts genome-scale gene expression profiling in two experimental systems that will enable a comparative genomics approach. The first system is maize hybrids that overexpress Glossy15, which exhibit delayed shoot maturation and accumulate

greater stem biomass and sugars compared to near-isogenic grain hybrids. The second system is a Saccharum pedigree where progeny vary for their relative production of stem sugar, fiber, and biomass. In both systems, the differences in compositional, anatomical, and gene expression differences are compared between the high sugar type and the high fiber type.

To Mom and Dad

ACKNOWLEDGEMENTS

I would like to thank many people for offering me support and guidance throughout this project. I would like to thank my advisor, Stephen Moose, for helping me write and make sense of my data. Thanks to the Energy Biosciences Institute and the Department of Energy for providing funding for my projects and to Anna Hale (USDA) for providing me with my *Saccharum* cane samples. I also appreciate Stefan Bauer at UC Berkeley for determining the chemical composition of my stem samples. Thank you to Kankshita Swaminathan for answering so many of my questions and helping me solve problems and to my other lab mates who helped me complete this work. I would also like to thank my undergraduate researchers, Jessica Mattick and Eric Dyer, for their assistance throughout my project. And finally, thanks to my family and friends for supporting me and encouraging me not to quit during difficult times.

TABLE OF CONTENTS

CHAPTER I
GLOSSY15 OVEREXPRESSION IN HYBRID MAIZE CAUSES CHANGES IN CARBON PARTITIONING

1.1 ABSTRACT

The *glossy15* (*gl15*) gene is an APETALA2-like gene that has been demonstrated to promote the juvenile phase in maize. The opposing actions of *gl15* and miR172 regulate the vegetative phase change. A previous study showed that increasing gl15 activity in maize increased leaves that express juvenile traits and delays flowering. We show here that the same increase in *gl15* causes this delay in flowering by repressing the expression of *FLOWERING LOCUS T* (*FT*), a known promoter of flowering in *Arabidopsis*. The delayed flowering also alters the source and sink interactions of the maize stem and ear causing an increase in sugars in the stem. This stem sugar increase makes the transgenic *gl15* plants a better feedstock for maize stover ethanol production.

1.2 INTRODUCTION

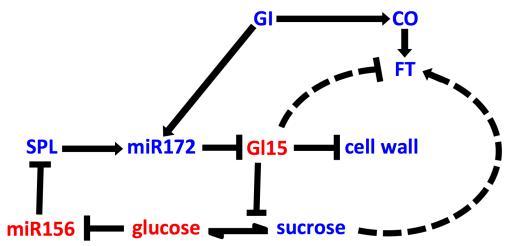
Due to an increase in worldwide population and higher demand from developing nations, oil prices have continued to increase in recent years. Fossil fuels are finite and new renewable, sustainable fuel sources are needed. According to U.S. Department of Energy Bioenergy Technologies Office, corn is the leading feedstock for producing ethanol fuel in the United States. Ethanol is produced from both kernels and stover. Starch much first be broken down into simple sugars before being converted to ethanol. Increasing the simple sugar content in maize stems could increase the ease of converting the stover to ethanol and in turn decrease the price.

Creating a maize that flowers later could cause an increase in stem sugar, but delaying flowering is not as simple as looking at a single trait. Multiple genes control this trait and the shoot development pathways are not fully understood yet. As plants

develop, they go through developmental phase changes. An important phase for biomass accumulation is the timing of the floral transition or the change from the vegetative phase to the reproductive phase (Lawson & Poethig, 1995; Kerstetter and Poething, 1998). The vegetative phase can be further divided into the juvenile phase and the adult phase. The juvenile phase is the time in plant development soon after germination where the plant develops true leaves and axillaries buds. The plant is incapable of sexual reproduction at this time. A plant is considered in the adult phase when it is able to reproduce sexually under normal conditions but does not have actual reproductive structures present (Poethig, 2003). These two phases can be identified by morphological traits but they vary among plants (Goebel, 1900; Allsopp 1967). The timing of these changes can effect how plants partition their carbon.

Figure 1.1: MircroRNA156/172 flowering pathway and genes. Panel A illustrates the part of the flowering gene pathway that is proposed for maize while panel B shows the number of paralogs for each member in the pathway in arabidopsis, rice, sorghum, and maize. The maximum likelihood tree in panel C was constructed using the related AP2 gene protein sequences. The test of phylogeny used was the Bootstrap method with 1000 replications and the heuristic method used was Nearest-Neighbor-Interchange.

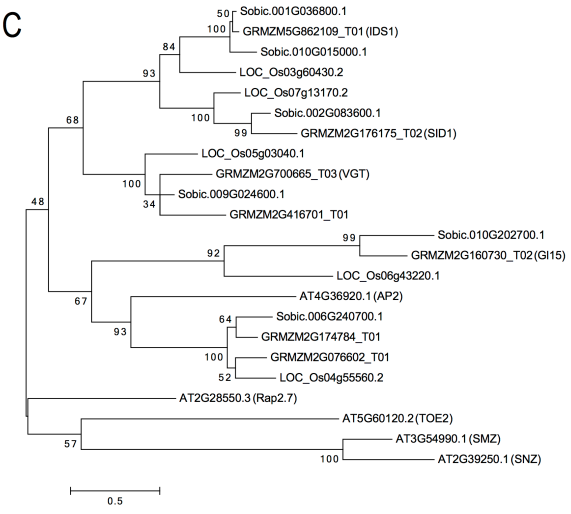
A



B

	Arabidopsis	Rice	Sorghum	Maize
miR156	8	11	9	11
SPL	8	10	9	16
miR172	5	4	6	4
AP2	5	5	6	7
FT	6	19	19	27
GI	1	1	1	2
CO	6	6	6	6

C



A key pathway in shoot maturation involves two microRNAs: *mircoRNA156* (*miR156*) and *mircoRNA172* (*miR172*) (Figure 1.1). These genes help regulate the transitions from the juvenile phase to the adult vegetative phase and on into the reproductive phase. The timing of these transitions is important for all plants. If a plant flowers before the growing season is over, it will produce less biomass because it will spend less time in the vegetative stage producing leaves. The targets of *miR156* include *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) genes in *Arabidopsis*. Of these genes, *SPL3*, *SPL4*, and *SPL5* are known to promote flowering. They each have a microRNA recognition element (MRE) for *miR156* in their 3' UTRs (Cardon et al., 1997; Gandikota et al., 2007; Wang et al., 2009; Wu et al., 2009; Wu and Poethig, 2006). Further downstream in the pathway, *SPL* genes promote *miR172*. As plants age, *miR172* levels increase as *miR156* levels decrease in the leaves. However, *miR172* levels are thought to be control by photoperiod rather than age (Jung et al., 2007). *miR172* promotes flowering by targeting transcription factors that repress flowering such as *APETALA2* (*AP2*), *SCHLAFMUTZE* (*SMZ*), *SNZ*, and *TARGET OF EAT* (*TOE*) genes (Aukerman and Sakai 2003; Chen 2004; Mathieu et al., 2009; Schmid et al., 2003). These AP2 transcription factors repress *FLOWERING LOCUS T* (*FT*), the hormone responsible for trigging flowering in plants. It has also been proposed that *miR172* is upregulated by *GIGANTEA* (GI) (Jung et al. 2007). *GI* also promotes flowering through an alternate path where it upregulates *CONSTANS* (*CO*) which then upregulates *FT* (Valverde et al., 2004; Mizoguchi et al., 2005).

Altering the expression of genes in this pathway can alter the way the plant traverses its lifecycle. Artificial elevation of *miR156* levels prolongs the juvenile vegetative phase. A delay in flowering has been observed in transgenic *Arabidopsis* that constitutively overexpress *miR156* (Schwab et al., 2005). These plants produce increased numbers of juvenile leaves and show reduced apical dominance. Not only is *miR156* key in promoting the juvenile phase, it is necessary for expression of juvenile traits. When mimicry targets sequester *miR156*, transgenic plants produced few leaves with only adult features and flowered soon after (Franco-Zorrilla et al., 2007; Todesco et al., 2010). As stated earlier, *miR156* levels decline over time based on age-dependent

factors. These factors are still unknown as we do not know how plants determine age. However, a recent study suggests that a factor that represses *miR156* is located in the leaves (Yang et al., 2011). It has also been suggested that sugar content plays a role in repressing *miR156* and promoting vegetative phase change. Higher levels of glucose in the plant cause a signaling pathway activated through *HEXOKINASE1*, which leads to repression of the microRNA (Yang et al., 2013).

Constitutive overexpression of *miR156*-resistant *SPL3*, *SPL4*, and *SPL5* in *Arabidopsis* has produced plants that flower and produce leaves with adult traits earlier than their wild counterparts confirming that *SPL* genes promote the adult phase. However, transgenic plants that overexpress *SPL3* only flower slightly earlier suggesting that *miR156* is very effective at repressing *SPL* genes (Cardon et al., 1997; Gandikota et al., 2007; Wang et al., 2009; Wu et al., 2009; Wu and Poethig, 2006). Similarly, overexpression of these targets and *miR172* mimicry targets in transgenic plants caused late flowering and overexpression of *miR172* caused early flowering (Aukerman and Sakai, 2003; Chen, 2004; Jung et al., 2007; Mathieu et al., 2009; Schmid et al., 2003).

Most of the studies involving the *miR156/miR172* pathway were performed using *Arabidopsis*, but there have also been studies in maize. The *Corngrass1* (*Cg1*) mutant shows increased tillering that leads to higher biomass and also has less lignin and more glucose in the leaves. This mutation is caused by a tandem *miR156* gene (Chuck et al. 2007). A similar delay in development occurred in switchgrass when transgenic plants are constructed using the maize *Cg1* gene (Chuck et al. 2011).

Other genes in the pathway can also be altered to produce a delay in flowering. *Glossy15* (*gl15*) is an AP2-domain containing gene that has been shown to regulate juvenile epidermal cell traits in leaves of maize. The gene gets its name from the glossy appearance of epicuticular wax on juvenile leaves. Loss-of-function mutations at the *gl15* locus have shown shorter expression of juvenile traits in leaves (Moose and Sisco, 1994). This gene was also shown to encode a putative transcription factor that shared

significant sequence similarity to AP2 and *AINTEGUMENTA* (*ANT*) in Arabidopsis (Moose and Sisco, 1996). Later studies showed that *miR172* represses *gl15* in maize and suggested that the *gl15* gene only can repress flowering in maize (Lauter, et al., 2005). The overexpression of *gl15* leads to the same suite of phenotypic changes seen in photoperiod-sensitive maize. One consequence of the delay in shoot maturation conditioned by *gl15* overexpression is a decrease in grain yield and increase in stover (Pulam 2012).

Studies have also looked into the source of florigen in maize. *FT* in *Arabidopsis* contains a phosphatidylethanolamine binding (PEBP) domain. In rice, two paralogous genes that are orthologous to *FT* are responsible floral activators. *Heading Date 3a* (*Hd3a*) is initiated under short days and *Rice flowering locus T1* (*RFT1*) was identified as a major floral activator under long day conditions (Kojima et al., 2002; Komiya et al., 2008, 2009). The PEBP family is large in maize and members are named *Zea mays CENTRORADIALIS* (*ZCN*). *ZCN15* maps to a region syntenic to the region where *Hd3a* is located in the rice genome but it was discounted as the source of florigen because it is expressed mainly in the kernels. Under ectopic expression, *ZCN8* was found to induce early flowering in transgenic plants and is considered the maize *FT* (Danilevskaya et al., 2008; Meng et al., 2011).

The availability of near-isogenic genotypes that differ only for *gl15* activity, coupled with the extensive resources for maize functional genomics, offers a simple and powerful system to directly assess the impact of changes in developmental timing on stem carbon partitioning. This study shows that the delayed flowering caused by the overexpression of this gene alters the sugar pathway in the stem. Also, the findings also suggest that overexpression of *gl15* directly or indirectly represses the expression of *Flowering Locus T* (*FT*).

1.3 MATERIALS AND METHODS

1.3.1 Plant Material

Stem samples collected for this study were taken from the maize hybrid B73xMo17, also referred to as control, and its transgenic counterpart that overexpresses the *glossy15* gene, also referred to as Gl15-TG. The plants were grown in Urbana, Illinois at the University of Illinois South Farms from May until August of 2011 and 2012.

1.3.2 Anatomy Sections

Every two weeks from July to September of 2011, samples were cut from the middles of stem internodes from each variety, locked into macro-cassettes (Electron Microscopy Sciences catalog # 70077-W), and placed in an amber container of 4% paraformaldehyde. 4% paraformaldehyde was made from 16% Paraformaldehyde (formaldehyde) aqueous solution (Electron Microscopy Sciences catalog # 30525-89-4), 10X PBS Buffer (VWR catalog # 16777-250), and ddH_20. Two replicates were taken for each variety per time point. The samples taken were no greater than one inch and were collected from every other internode along the stem starting at the bottom most internode.

The first step to fixing the samples was to place the uncapped sample collect bottle in a desiccator (Fischer Scientific catalog # 08-642-7) for 24 hours under a vacuum. After this process, the cassettes were transferred to a Leica ASP300 3 (Leica Biosystems) automated vacuum tissue processor. Once in the machine, the samples were taken through a 13-step program were they were soaked for five hours in the following: neutral buffered formalin, 70% ethanol, 70% ethanol, 80% ethanol, 95% ethanol, 100% ethanol, 100% ethanol, 100% xylene, 100% xylene, paraffin wax at 58°C, paraffin wax at 58°C, paraffin wax at 58°C. Each step lasted for five hours. The samples were embedded using a heated paraffin embedding module Leica EG1150 H

(Leica Biosystems). The wax used for infiltration and embedding was Richard-Allan Scientific Paraffin Type H (Fischer Scientific catalog # 22-050-128). Embedded sample blocks were left to harden overnight. Sections were cut to 8 µm using a Leica RM2255 microtome (Leica Biosystems) with a low profile diamond blade (Sturkey catalog # DT315D50) and placed on Superfrost Excell microscope slides (Fischer Scientific catalog # 22-037-247). Slides were left overnight to dry.

Slides were then stained with Calcofluor White to look at cellulose and Schiff's reagent to look at lignin. Before staining, slides were baked at 65°C for one hour. Soaking for three minutes in the following rehydrated sections: 100% xylene, 100% xylene, 100% xylene, 100% ethanol, 95% ethanol, and ddH_2O. Slides were laid flat and drops of Schiff's Reagent (Sigma-Aldrich catalog # 3952016) were added until the stain covered the entire section. The slides were then covered and allowed to stain for 15 minutes. After staining, the slides were rinsed carefully with ddH_2O for five minutes. The Calcofluor White stain was then applied in a similar manner and left to stain for 15 minutes. Calcofluor White 100x solution is prepared by adding 3.5 mg of Fluorescent Brightener 28 (Sigma-Aldrich catalog # 3543) to 0.5 ml of ddH_2O, increasing the pH to 11, and adjusting the volume to 1 ml. The 100x solution is then diluted to 1x using Tris-HCL Buffer, 0.1M pH 9.0. After the allotted time, the slides should be rinsed with ddH_2O for three minutes changing water frequently as it turns pink and then placed in a 1X PBS bath until they are mounted. Slides were mounted using ProLong® Gold Antifade Mountant with DAPI (Life Technologies catalog # P-36931).

Stained stem sections were observed under a Zeiss LSM710 confocal microscope (Zeiss). Only sections of bottom ear internode were examined. The Schiff's Reagent was excited at 561 nm laser and the emission was observed in the 575 to 650 nm range while the Calcofluor White had an excitation wavelength of 405 nm and an emission spectrum of 411 to 460 nm.

1.3.3 Composition Analysis

During the 2012 growing season, samples were collected from the first ear internode of both varieties. The internode was divided in half longitudinally and half was used for composition analysis. The other half of the sample was used for RNA extraction. Four replicates were taken for each sample. Samples were collected on July 3rd, July 13th, July 26th, and August 9th for both varieties. An additional transgenic plant sample was taken on August 24th. These samples were stored on dry ice until immediately after being cut. The tissue was lyophilized for 48 hours using the automatic settings of the FreeZone 6 Freeze Dry System (Labconco).

Biomass was milled using a commercial electrical coffee mill and then homogenized by ball-milling for five minutes in a canister ball-mill (model 8200, Kleco, Visalia, CA). The biomass was vacuum-oven dried at 45°C for 15 hours and 300 mg of the dried biomass was extracted with 3x2 mL of 70°C water and 3x2 mL of 70°C ethanol in a syringe cartridge and then vacuum-dried at 45°C for 15 hours as described in (Kuchelmeister and Bauer, 2014). Extractives were calculated as the percentage weight loss based on the initial biomass used for extraction.

Soluble sugars (sucrose, glucose, fructose) in the water phase were quantified after spiking in about 1 mg/mL fucose as internal standard to determine the extraction volume and to compensate for sample dilution before analysis. The sugars were analyzed using an ICS-3000 HPLC system (Dionex) equipped with a pulsed-amperometric detector. Samples were injected onto a 150 x 3 mm i.d. PA20 column (Dionex) with a 50 mm x 3 mm guard column of the same material and eluted at 30°C with a mobile phase of 30 mM KOH at a flow rate of 0.4 mL/min.

About 50 mg of the extracted and vacuum-dried biomass were hydrolyzed by a one- (4%) and two-stage (72%/4%) sulfuric acid hydrolysis step for the release of monosaccharides as described in (Bauer and Ibáñez 2014). A sugar recovery standard (SRS) containing glucose, xylose and arabinose was prepared in the same acid

concentration used during autoclaving. One aliquot was kept at 4°C and another one with a volume of 14.5 mL was autoclaved (121°C) together with the samples (Sluiter et al., 2012). After autoclaving, mixtures were cooled, vortexed and stored at 4°C for 16 hours in order to precipitate insoluble material. A 0.5 mL aliquot of the clear supernatant was filtered (0.2 µm PES) and used for monosaccharide HPLC analysis.

For the determination of total glucose, xylose and arabinose content released, samples were injected onto an HPX-87H (300 x 7.8 mm, Bio Rad, Richmond, CA) column with a 30 x 4.8 mm cation H guard column (Bio Rad). The instrument (1200 series, Agilent Technologies, Santa Clara, CA) was equipped with a refractive index detector. Elution was performed at 50°C with 5 mM sulfuric acid at a flow rate of 0.6 mL/min. On the HPLC column used, galactose and mannose co-elute with xylose and these sugars were therefore quantified together as parameter “xylan” during total sugar analysis. For calculation of polysaccharides content, conversion factors of 0.9 (glucan) and 0.88 (xylan, arabinan) were used. The difference of the total glucan content determined by the two-stage acid hydrolysis and the glucan content determined by the one-stage acid hydrolysis was referred to as “cellulosic glucan (cellulose)”.

Precipitated solids of the two-stage sulfuric acid hydrolysis were resuspended and the suspension filtered through previously pretreated (575°C for 3 hours) ceramic filter crucibles (pore size M, Willmad LabGlass, Vineland, NJ, USA), or previously pretreated (575°C for 3 hours) glass microfiber filters (AP40, 25 mm, EMD Millipore™, Billerica, MA, USA) using a filtration flask and reduced pressure, and all retained solids were washed extensively with deionized water (Ibáñez and Bauer, 2014). The weight of residue and crucibles or glass microfibers was determined after drying at 105°C for 16 hours and also after subsequent ashing at 575°C for 3 hours. The difference of these two weight determinations was referred to as Klason lignin (ash corrected).

1.3.4 RNA Extraction and Sequencing

The samples used for sequencing were the other half of the longitudinal sliced ear internode samples used for composition analysis. These tissues were ground with liquid nitrogen using a combination of A 11 basic analytical mills (IKA catalog # 0002900001) and mortars and pestles. Total RNA was isolated using the protocol developed for pine (Chang et al., 1993) with some modifications. CTAB buffer was prepared without the addition of spermidine. The buffer was not heated and equal volumes of CTAB buffer and acidified phenol pH 4.3 ± 0.2 with isoamylalcohol were added to each tube prior to the addition of the samples. The tubes used for the phase separation steps were 5 PRIME™ Phase Lock Gel™ (Fischer Scientific catalog # FP2302820) so centrifugation speeds were adjusted accordingly and shaking by hand replaced vortexing for these steps. The lithium chloride step was replaced with an isopropanol precipitation followed by 75% ethanol wash. Samples were then treated with DNase, reprecipated with isopropanol and washed with 75% ethanol. Pellets were dried using a Savant™ DNA SpeedVac™ Concentrator (Thermo Scientific) and resuspended in water. The yield of the mRNA was quantified with a NanoDrop Spectrophotometer ND-100 (Thermo Scientific). Quality was checked using Agilent 2100 Bioanalyzer (Agilent) and only samples with a RNA integrity number of 8 or higher were used for sequencing.

Libraries were prepared with Illumina's 'TruSeq RNAseq Sample Prep kit' by the W.M. Keck Center at the University of Illinois. Combining one replicate from each condition generated four pools and the pooled libraries were quantitated by qPCR and then sequenced. There were two types of reads generated: mRNA, 100nt length reads and small RNA, 50nt length reads. The mRNA libraries were sequenced on one lane for 101 cycles from one end of the fragments. The small RNA libraries were sequenced on two lanes for 51 cycles. Both runs were performed on a HiSeq2000 using a TruSeq SBS sequencing kit version 3 and processed with Casava 1.8.2.

1.3.5 mRNA-Seq Analysis

A total of 841 million reads were generated from all the libraries. The mRNA reads were trimmed for quality and adaptors using the single end mode of Trimmomatic-0.30 (Bolger et al., 2014). An individual adaptor file was generated for each library based on the adaptor used. Bases were cut off the end of the read if the phred score was below 30 and any read less 40 nt was discarded. On average, 99% of the reads met the quality threshold.

Reads were aligned using TopHat v2.0.8 (Kim et al., 2013) to the B73 genome v2 with the aid of the 5b.60 annotation (http://www.phytozome.net/maize.php; Schnable et al., 2009). Reads were only aligned to chromosomes. Default parameters were used unless stated and the following parameters were changed in an attempt to capture the known allelic divergence between the hybrid and the B73 genome: read-mismatches 5 (95% identity), read-gap-length 3, read-edit-dist 5, min-intron-length 25, max-insertion-length 5, max-deletion-length 5, max-multihits 10, microexon-search, b2-very-sensitive. The average alignment rate was 93% with 79% of those mapping having a unique alignment. The resulting BAM file was then converted to SAM format and passed to HTSeq (Anders et al., 2014) to obtain read counts. HTSeq counted uniquely mapped reads with a mapping quality score of 50 at the gene level using the intersection_nonempty mode with gene coordinates based on the 5b.60 GFF3. The resulting libraries averaged 75% of the total trimmed reads.

Read counts were examined using five Bioconductor packages: affycoretools (MacDonald, 2008), WGCNA (Langfelder and Horvath, 2008, 2012), limma (Smyth, 2005), and dendextend (Galili, 2014); and differential expression was determined using the edgeR package (Robinson and Smyth, 2007, 2008; Robinson et al., 2010; Robinson and Oshlack, 2010; McCarthy et al., 2012). Of the 39,423 B73 genome v2 5b.60 chromosomal gene models considered, 35,207 had at least one read that mapped in a sample. Counts per million were calculated and a gene was kept if there was at least one day where one treatment showed an average of 5 counts per million or higher. The

threshold of 5 counts per million was chosen because it created a normal distribution of the reads and this filtering step left 20,583 gene models and 99% of the reads in all libraries (Figure 1.2). Based on the hierarchical clustering analysis and PCA, two replicates were discarded from the control, July 3rd group and one replicate was discarded from every other group. A contrast matrix was designed to compare GI15-TG versus control by sampling date. Using the trimmed mean of M-values (TMM) method further normalized the libraries and improved their clustering (Figure 1.3). Variance among replicates was calculated for each gene using the edgeR tagwise estimate and the glmFit function was used to model the data to a generalized linear model. The p-values were calculated using likelihood ratio test for the coefficients in each row of the contrast matrix where the null hypothesis states that there is no difference in coefficients. Genes were considered differentially expressed if they had a false discovery rate equal to or less than 0.001 and a fold change greater than or equal to 4 where fold change and counts per million thresholds were determined by examining MA plots (Figure 1.4). The first filtering (Figure 1.2) included every sample disregarding sampling day when considering if the gene met the filtering criteria. This left some genes that were expressed at low levels in the individual sampling day comparisons so each gene set was further filtered to leave only genes with five counts per million or higher in one of the treatments on that particular sampling day.

Figure 1.2: Filtering of Raw Reads. The filter was implemented by averaging the replicates for a condition. Genes passed the filter if the replicates of at least one variety had an average of 5 counts per million or higher on at least one day. Panel A shows all of the genes before filtering. The high peaks on the left show that many genes have low expression. These low expressed genes are filtered out in panel B leaving the data with a nearly normal distribution.

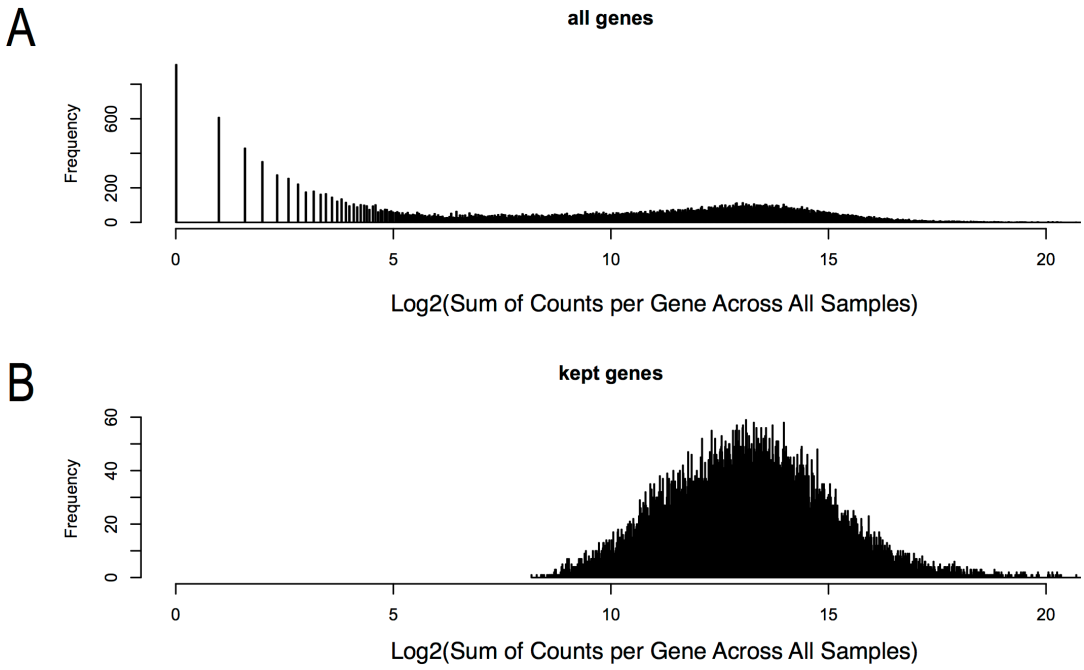


Figure 1.3: Removal of some replicates greatly improves clustering. Panels A and B show principal component analyses of the five counts per million filtered and TMM normalized libraries. Panel A shows all of the sample replicates and panel B shows the improved clustering after removal of replicates. The hierarchical clustering analysis in panel C also confirms that removing these replicates reduces variance in the experiment. The main sources of variance in the experiment appear to be age of stem and genotype as seen by the principal component analysis.

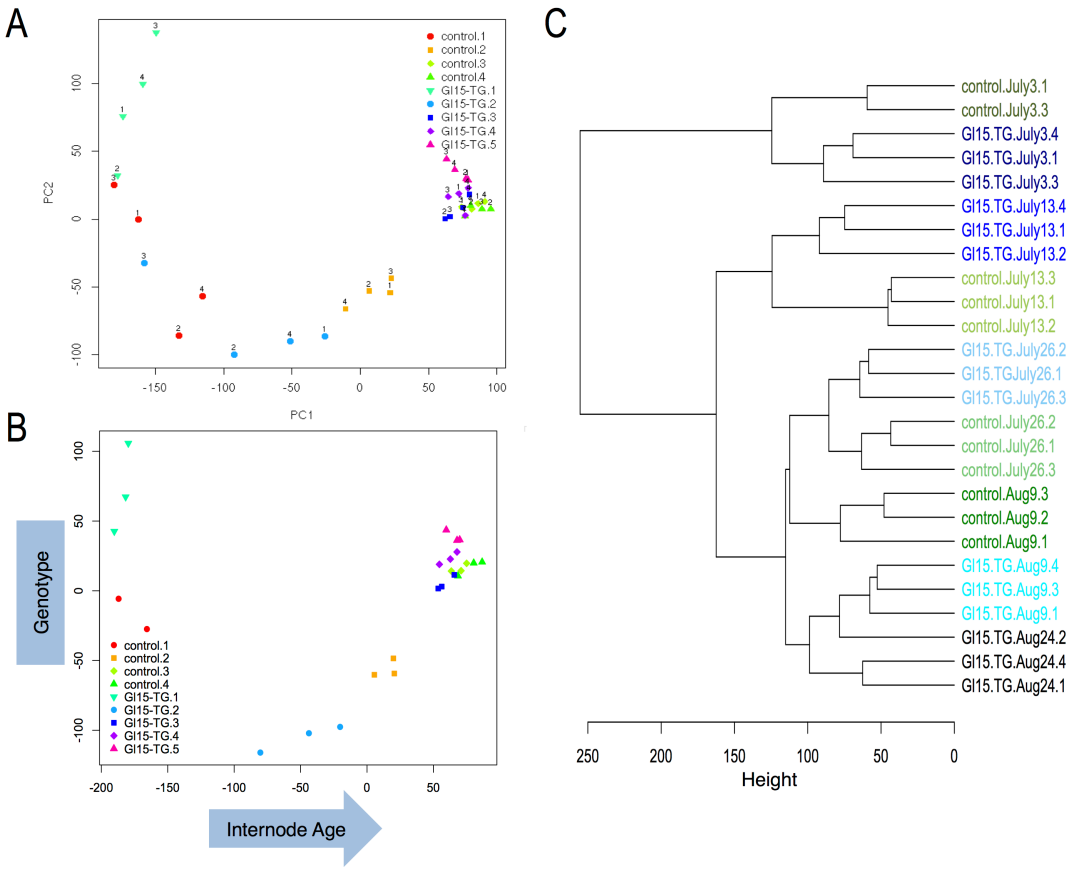
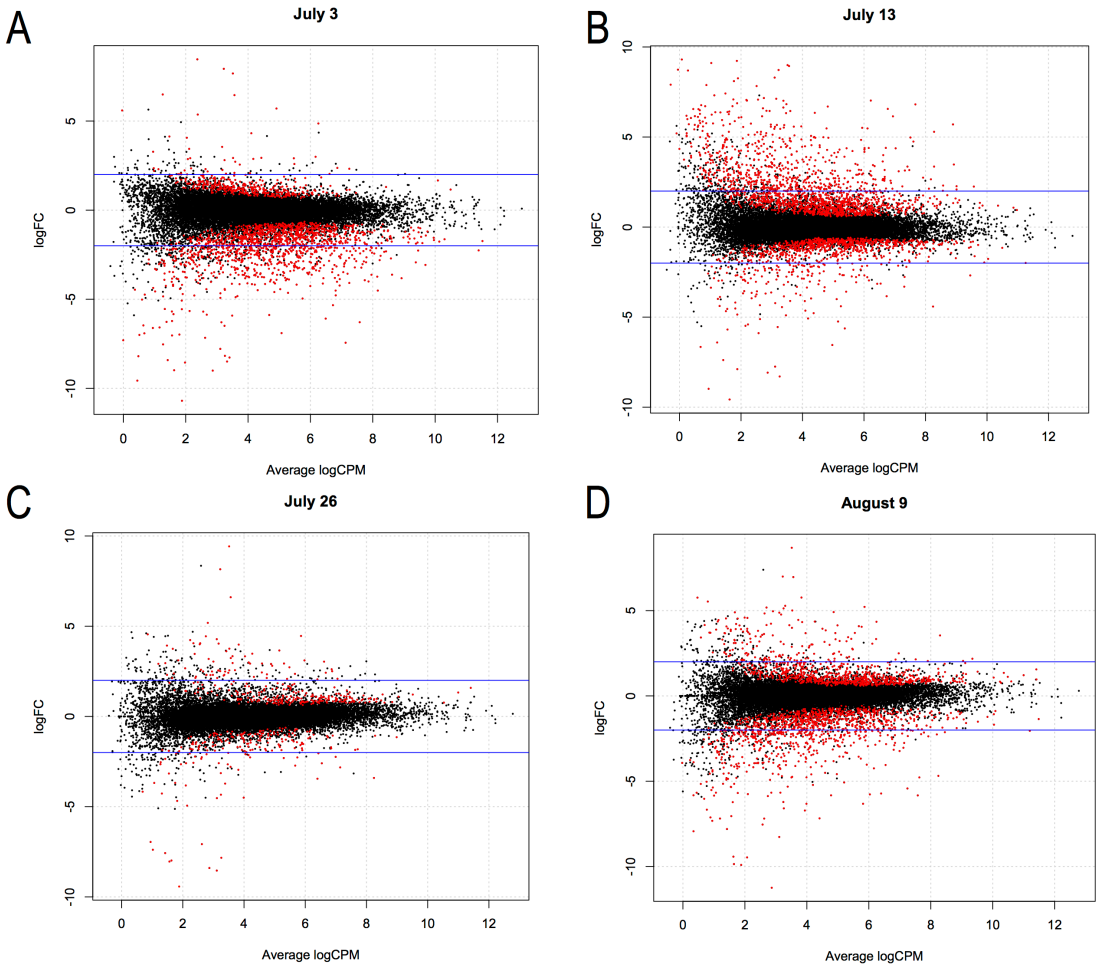


Figure 1.4: MA plots by sampling date where both varieties were sampled. The red dots indicate genes that have a false discovery rate less than or equal to 0.001 and the blue lines indicate 2 and -2 log base two fold change. Panels A, B, C, D and correspond to the four sampling date comparisons between the transgenic and control.



1.3.6 miRNA-Seq

The small RNA sequencing generated 765 million reads. These reads were aligned to the same version of the B73 genome as the mRNA reads. The alignments were performed using the Novoalign program (Novoalign v3.00.02 http://www.novocraft.com/main/index.php). Adaptors were trimmed by the program, detection of hairpin scores was turned on, minimum length was set to 18 nt, the score threshold was at 30, the homopolymer filter score was set at 90, and all alignments

meeting these criteria were reported. On average, 90% of the reads met the quality threshold parameters and 52% of the reads aligned. Read counts were obtained using HTSeq in a similar manner to the mRNA reads. The same parameters were used but the maize GFF3 for miRNA was downloaded from miRBase (ftp://mirbase.org/pub/mirbase/CURRENT/genomes/zma.gff3). This GFF3 file corresponds to the B73 v2 genome that was used for alignments. The resulting library sizes accounted for 0.05% of the total mapped reads.

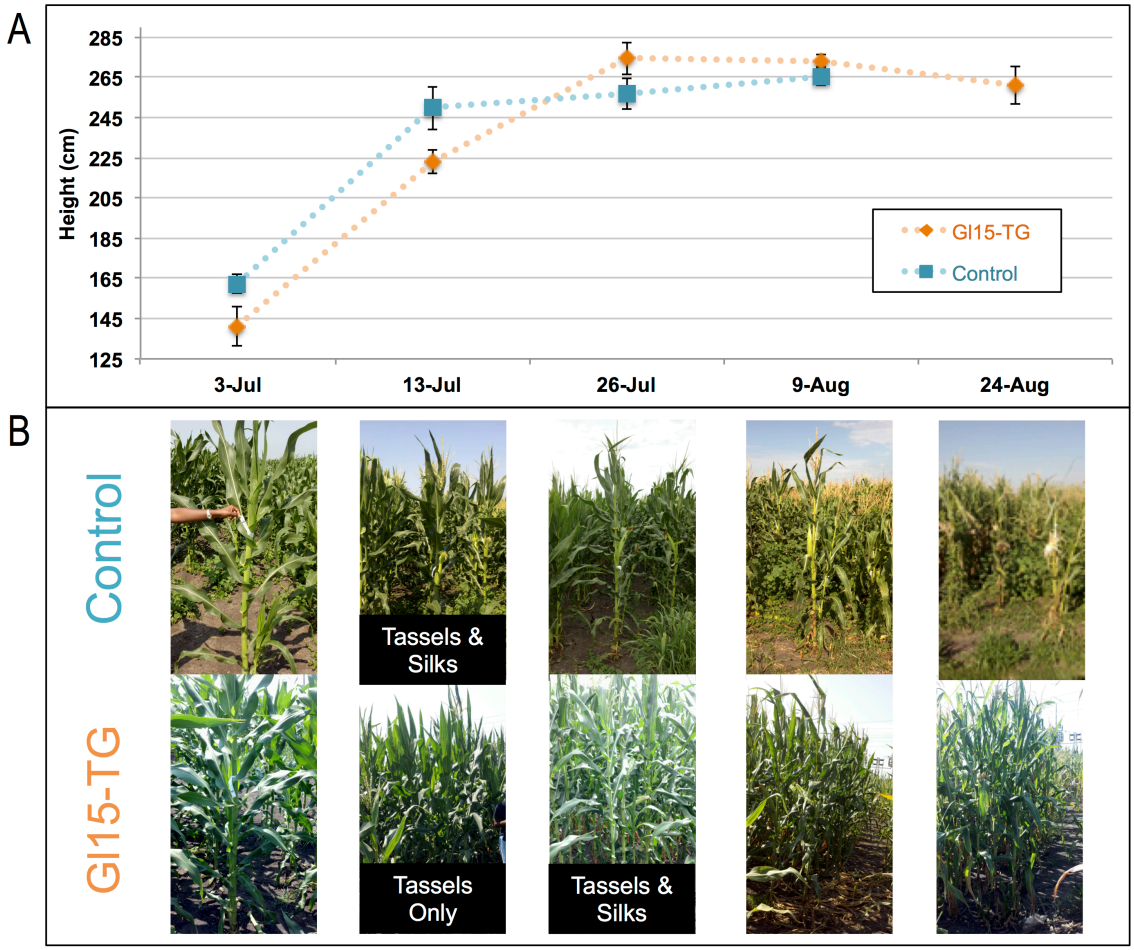
The same edgeR pipeline used to analysis the mRNA sequencing was attempted on this data; however, the library sizes were not consistent, possibly due to miRNA cloning bias. Instead, the replicates for each treatment by sampling day were averaged. The raw count value was transformed by taking the logarithm base 2. These values were then imported in R and a heatmap was constructed using the Heatplus package (Ploner, 2014).

1.3.7 miRNA qPCR Analysis

Expression of three miRNA, miRNA 156, miRNA 172, and miRNA 168, was quantified using Applied Biosystems® TaqMan® MicroRNA Reverse Transcription Kit and the Applied Biosystems® TaqMan® MicroRNA Assay (Applied Biosystems). Total RNA from the samples used in the sequencing assay was reverse transcribed and the cDNA obtained was analyzed on the Lightcycler® 480 System (Roche) using Applied Biosystems® TaqMan® Universal PCR Master Mix, No AmpErase® UNG (Applied Biosystems). The protocols followed were described in the assay kits. Raw data was exported from the machine and imported into Microsoft Excel where 483-533 values were used to calculate second derivatives. This data was then exported to a Perl script where the highest second derivative value was used to pick the cycle threshold (C_T) for each sample. All four biological replicates were run and all samples taken on the same day were run on the same plate. The expression of miRNA 168 was used to normalize expression across samples so three technical replicates of this miRNA were performed for every sample. Expression of miRNA 172 and miRNA 156 was calculated as 2^(miRNA 168 C_T)/ 2^(target C_T).

1.4 Results

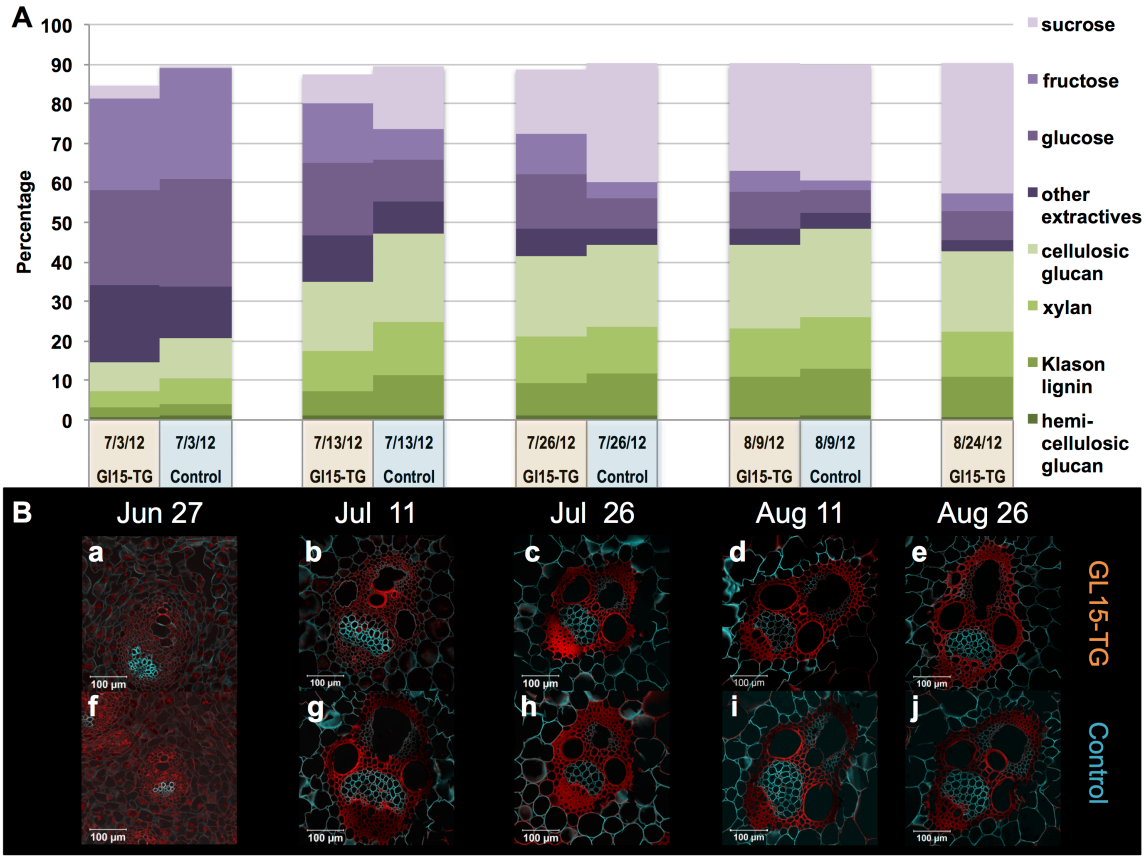
Figure 1.5: Growth of transgenic and control plants. The transgenic plant shows a delay in development and a later flowering date than the control. Due to their later flowering, the transgenics also grow taller than the controls.



The increased activity of *gl15* in the transgenic plant delays reproductive development in a similar way as described in Lauter et al., 2005. The control plants are taller than the transgenics pre-flowering. After the control plants flower, the transgenics continue to grow and eventually over take them. On July 13th, both tassels and silks were present on the controls while only tassels were present on the transgenics. The transgenic silks had emerged by July 26th. Post flowering, there was also a delay of

senescence with the trangenics remaining green longer than the controls. Over all, there seemed to be a ten day delay in development (Figure 1.5).

Figure 1.6: Compositional Comparison of Bottom Ear Internodes in GI15-TG vs Control. In panel A, the compositions of stems collected over the summer of 2012 are shown. The shades of purple indicate sugar types and the shades of green indicate fiber types. Panel B displays the anatomy of growing stems of both lines. The red indicates areas stained with Schiff's reagent and the blue shows areas stained by Calcofluor White.



The Schiff's reagent and Calcofluor White staining did not show a visible difference between transgenic and control stems. However, the compositional analysis showed that as the stems develop, they do have different carbon content (Figure 2.6). On July 3rd, neither of the plants was flowering and sugar accounted for 69% of the

composition in both types with the transgenic having slightly more sucrose and other extractives and the control having more glucose, fructose, and stem fiber components.

By July 13th, the control plants were fully flowering while the transgenics only showed tassels present. Flowering altered the composition of the stems with sugar only accounting for 42% of composition; however, fiber rose 47%. The predominant fiber type in the control stems is cellulosic glucan and sucrose became the predominant sugar. The fiber and sugar content of the transgenic stems were only 35% and 53% respectively. Cellulosic glucan was also the main source of fiber but the sucrose content was less than half of the control plants. Most of the sugar in the stem was in the form of glucose or fructose and the control plants still showed higher percentages of all fiber components.

Both tassels and silks were present on the transgenic plants on July 26th and there was now only a much smaller difference in stem sugar and fiber percentages between the two lines. However, the sugar content of the two stems were still not distributed similarly. The control plants had about twice the amount of sucrose compared to the transgenics. Sucrose had now also become the predominant sugar type in the transgenics but glucose and fructose still accounted for a large portion of total sugar content. The sucrose content in the transgenic stem on this date is very similar to the amount in the control stem on July 13th.

Samples taken August 9th show that the control is very similar to the sample taken on July 26th. The transgenic stem had also started to accumulate sucrose at a similar level as the control, but the transgenic still had overall higher sugar content due to elevated levels of glucose and fructose. Only the transgenic plant was sampled on August 24th due to the senescence of the control plants. On this final day, the transgenic showed about the same percentages of fiber as the previous collection date but had converted more glucose and fructose into sucrose (Figure 1.6).

Figure 1.7: DE Genes by date between Gl15-TG vs Control. Genes are significant at: false discover rate pvalue <= 0.001, Max Average counts per million >= 5, and fold change>=4. The bars in panel A indicate the number of genes that were higher between the two lines. July 3rd and August 9th both show the control having more genes with higher expression while July 26th the expression is about equal. July 13th is the only day sampled where there is a significant number of genes that are more highly expressed in the transgenic. Panel B shows the overlap of differential expression by sampling date. There are 1,406 genes differentially expressed of the 39,423 chromosomal gene models. Of these 1,406, 14 were differentially expressed on all dates sampled.

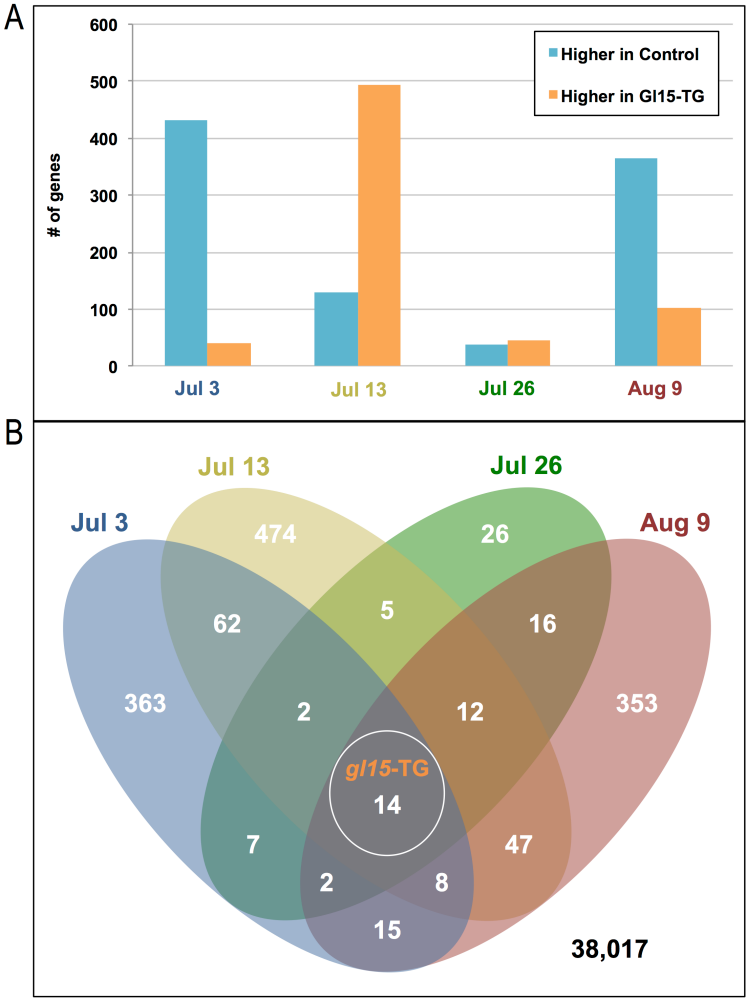
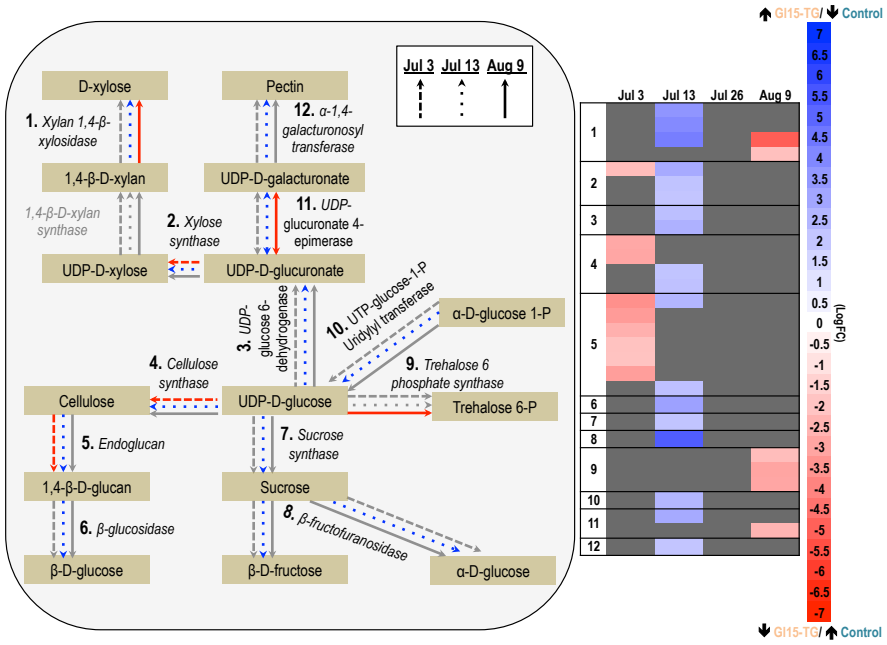


Table 1.1: Genes differentially expressed on all sampling dates. The bolded line marks *glossy15*, GRMZM2G160730. The average fold change shows the fold change or negative inverse fold change for numbers less than one that was calculated by averaging the counts per million from all of the transgenic samples and dividing by the average of the control samples. The chromosome location of these genes shows many are located on chromosomes two and eight. It is important to note that the location given for *glossy15* is the native gene location and not the transgenic.

Maize Gene ID	*Arabidopsis thaliana* Annotation	*Oryza sativa* Annotation	Average Fold Change	Chromosome Location
GRMZM2G457697	enhancer of rudimentary protein, putative	enhancer of rudimentary protein, putative, expressed	7	2: 32266968 - 32270914
GRMZM2G348664			164	2: 37244019 - 37246723
GRMZM2G148147			302	2: 43000189 - 43001179
GRMZM2G140160	RING/U-box superfamily protein	zinc finger, C3HC4 type domain containing protein, expressed	-286	2: 183406659 - 183415371
GRMZM2G082296	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	mRNA cap guanine-N7 methyltransferase 2, putative, expressed	15	5: 14523218 - 14525382
GRMZM2G071808		expressed protein	-547	7: 63920174 - 63924135
GRMZM2G587368		expressed protein	19	8: 16545140 - 16549180
GRMZM2G031917			-18	8: 21810058 - 21815510
GRMZM2G113415	CAP-binding protein 20	RNA recognition motif containing protein, putative, expressed	-18	8: 134165416 - 134166710
GRMZM5G804941			-1918	8: 140393076 - 140393403
GRMZM2G126917			13	8: 149218592 - 149219078
GRMZM5G878615	actin 3	actin, putative, expressed	-191	8: 163242789 - 163247999
GRMZM2G122057			-10	8: 164078649 - 164080712
GRMZM2G160730	**Integrase-type DNA-binding superfamily protein**	**AP2 domain containing protein, expressed**	**553**	**9: 95739338 - 95742681***

The results from the differential expression analysis of the RNAseq data show that on July 3rd and August 9th the majority of the differentially expressed genes, 91.3% and 77.9% respectively, were expressed at higher levels in the control plants. However, on July 13th, the majority, 79.2%, were shown to be expressed higher in the transgenic. Results from July 26th showed that the total difference in expression dropped to about one sixth that of the other days with an almost equal distribution of genes expressed highly in two lines. There were a total of 1,406 differentially expressed genes with many of them showing difference in expression on multiple days (Figure 1.7). There are only 14 genes that are differentially expressed at all sampling dates and among these genes is *glossy15*. Of the other 13 genes, six are higher expressed in the transgenic and seven are higher in the control. Four of these genes are located on chromosome 2 and seven are located on chromosome 8. Among these 14 genes only GRMZM2G457697 shows expression greater than 1.5 counts per million in the lower expressing line. GRMZM2G457697 is also the most highly expressed gene on the list (Table 1.1).

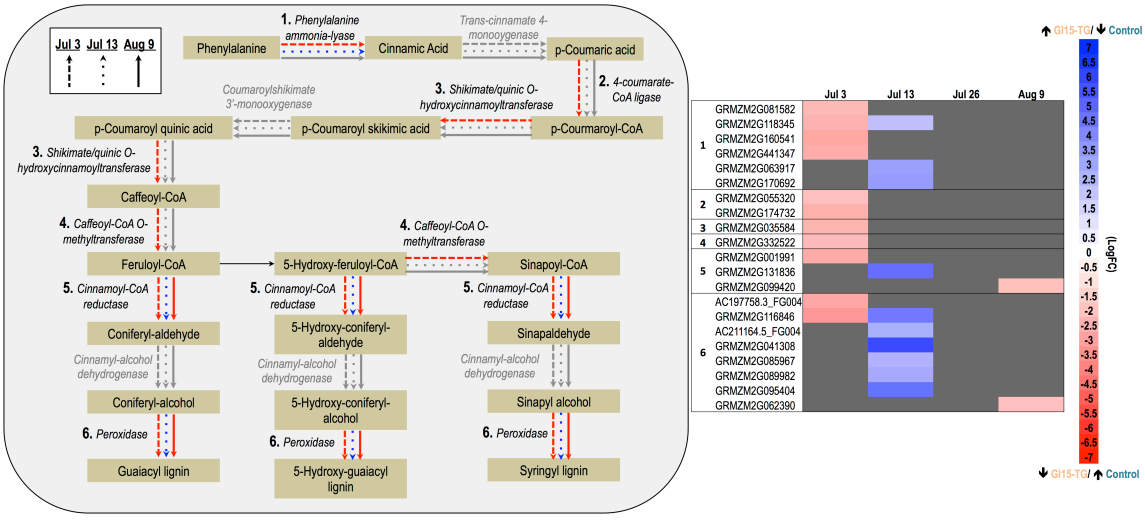
Figure 1.8: Effects of GI15-TG expression on Sugar Metabolism. The graph shows the pathway members that are affected by transgene expression over the four sampling dates. In the heatmap, blue indicates higher in transgenic/lower in control and read is the opposite. Dashed arrows, dotted, and solid arrows represent expression in July 3rd, July 13th and August 9th respectively. July 26th is not represented by any arrow because it showed no differential expression for genes in the pathway. Gray boxes and arrows indicate that the gene was not considered differentially expressed but my still have measured expression on that date. The number next to the enzyme correlates with the numbered boxes in the heatmap. Arrow directions and enzyme/substrate names are based on http://www.kegg.jp/kegg-bin/show_pathway?140425420526282/ko00500.args.



As seen in the composition analysis results (Figure 1.6), the addition of the transgene causes changes to the carbon partitioning in the stem. The transgenic matures slower and accumulates higher amounts of total sugar in the stem. When looking at gene expression of key enzymes in the sugar metabolism pathway, differences can be seen between the two lines as well. On July 3rd, the pathway only shows higher expression in the control plants for cell wall related enzymes. Cellulose

synthase and endoglucan build glucose into cellulose and there is an increase in the amount of cellulose in the control versus the transgenic on that day. Expression on July 13th shows a shift to all of the genes differentially expressed in the pathway being higher in the transgenic. This parallels the physiological delay in development caused by the transgenic. There appears to be a push among the genes to increase glucose in the stem and this is seen in the composition because glucose is the highest sugar component. Cellulose synthase and sucrose synthase are also pumped up on this day and there is an observable difference between the transgenic on July 3rd and July 13 for these components; however, there is also a higher expression of genes that break down these two into glucose. July 26th shows no difference is expression for the sugar metabolism pathway. This correlates with the similar composition profiles for that date. Without genes pushing the carbon toward glucose, more sucrose was produced but the amount lags behind the control stems. At this time, sucrose is produced for transport to the ear where it is converted to starch for seed development. On the final comparison date, August 9th, only a few genes are differentially expressed and they are again higher in the control. It is not certain how *xylan 1,4-β-xylosidase, trehalose 6 phosphate synthase,* and *UDP-glucuronate 4-epimerase* relate to the composition because xylan appears unchanged and pectin was not measured in our analysis. There is a reduction in glucose that could be in way related to the higher expression of *trehalose 6 phosphate synthase*.

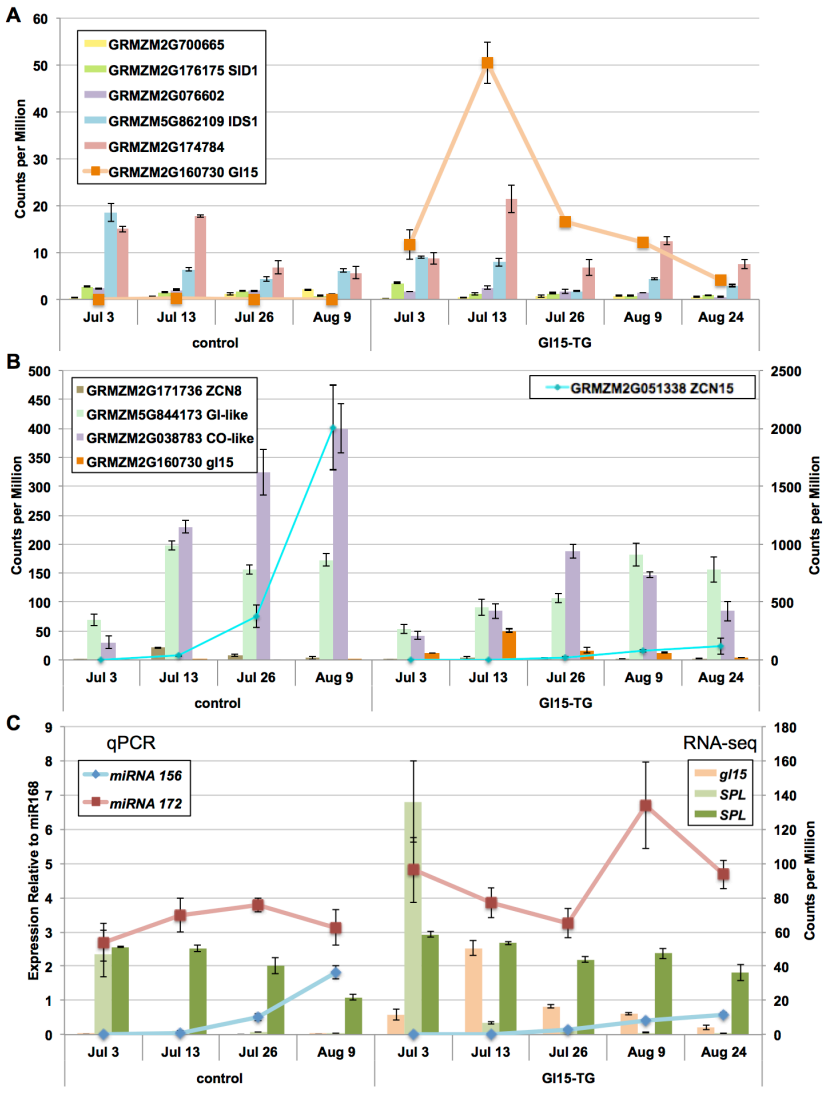
Figure 1.9: Effects of GI15-TG expression on Lignin Biosynthesis. The graph shows the pathway members that are affected by transgene expression over the four sampling dates. In the heatmap, blue indicates higher in transgenic/lower in control and read is the opposite. Dashed arrows, dotted, and solid arrows represent expression in July 3rd, July 13th and August 9th respectively. July 26th is not represented by any arrow because it showed no differential expression for genes in the pathway. Gray boxes and arrows indicate that the gene was not considered differentially expressed but my still have measured expression on that date. The number next to the enzyme correlates with the numbered boxes in the heatmap. Arrow directions and enzyme/substrate names are based on http://www.kegg.jp/kegg-bin/show_pathway?140425420526282/ko00500.args.



The lignin metabolic pathway shows a similar trend to the sugar metabolic pathway (Figure 1.8) with no differential expression on July 26th, genes only being expressed higher in the control on July 3rd and August 9th, and July 13th having only genes that are highly expressed in the transgenic. On July 3rd, there is a push in the control to produce more lignin. This correlates with the composition results (Figure 1.6) where the control has higher lignin content. July 26th shows the delayed response in the transgenic where genes are now being more highly expressed in those stems and more lignin is being produced. There is not a difference in expression on July 26th for lignin pathway in the two lines and the observed lignin percentages are equal, 11%.

However, on August 9th, there is once again a push in the control to produce more lignin and the composition shows a resulting increase in 2% more lignin in the control over the transgenic.

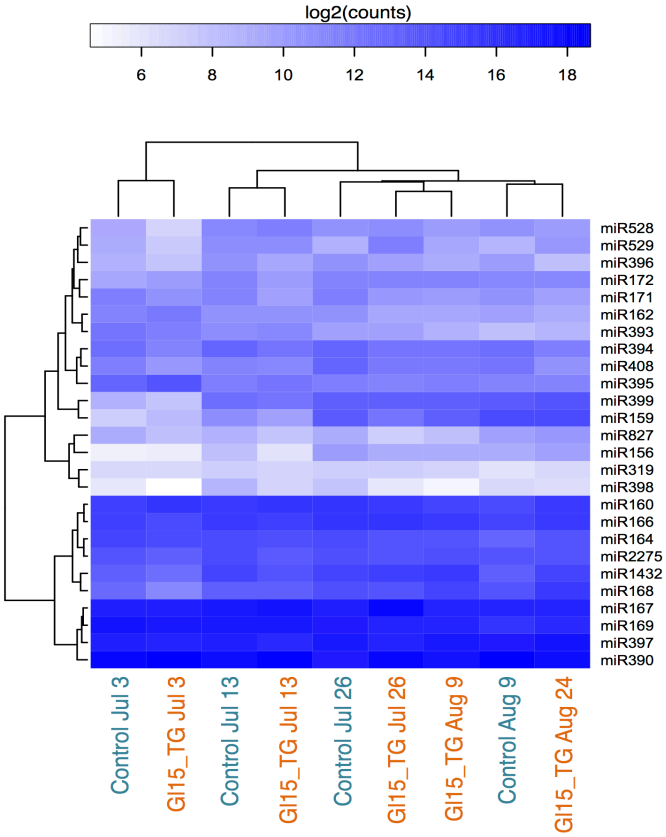
Figure 1.10: Effects of *gl15* Overexpression on other member of its regulation pathway. The expressions of maize AP2 genes that have *miR172* binding sites are shown in panel A. Panel B illustrates the difference in expression of flowering genes *FT-like ZCN genes*, *CO-like*, *GI-like* between the two lines and also shows *gl15* expression. Tassels and silks had emerged by July 13th for the control and July 26th for the transgenic. Panel C shows the expression of *miR156* and *miR172* on the left axis as well as two *SPL* genes and *gl15* on the right axis.



The RNASeq analysis showed that *gl15* was expressed at higher levels in all sampling dates (Figure 1.7) but this elevated expression did not significantly affect how other AP2 genes with miRNA172 binding sites were expressed. These results also showed that the expression of *FT-like genes ZCN8 and ZCN15*, on and after tassels and silks were observed in the control (July 13th), was 2 to 6 and15 to 25 fold higher respectively in the control plants than the transgenics. This was coupled with a similar suppression of *CONSTANS-like* (*CO-like*) on and after flowering and a slight suppression of *GIGANTEA* (*GI-like*). The levels of *GI-like* are recovered after the transgenic tassels and silks have both emerged on July 26th. The transgenic also showed a slight increase in *SPL* expression on July 3rd and August 9th (Figure 1.10).

The expression of miR172 and miR156 was quantified using qPCR. SPL genes upregulate miRNA172 and there is a correlation between the July 3rd and August 9th spike in *SPL* expression in the transgenic with a spike in miR172 expression on those dates. After the control flowers on July 13th, there is a progressive increase of miR156 expression in the control. There is a similar post flowering increase in the transgenic but it appears that the increase is repressed in the transgenics (Figure 1.10).

Figure 1.11: miRNAseq Heatmap. The log base 2 values were taken from the raw count data after averaging replicates. The clustering shows that internode age is still the major source of variance in the samples with the July 3rd samples being the most different from the other dates. The trend of miR156 increasing with age in the stems matches the results from the qPCR and also shows that miR156 is repressed in transgenics.



The miRNAseq heatmap analysis showed a similar clustering of samples as the RNAseq analysis (Figure 1.3) with samples clustering from youngest to oldest and July 3rd being the most different among all the dates. The miRNA clustered into three groups: highly expressed, moderately expressed, and faintly expressed; with miR156 being a part of the faintly expressed ground and miR172 being a part of the moderately expressed group. The progressive post flowering increase in miR156 expression and the repression in the trangenic that was seen in the qPCR is present in this analysis as

well. However, trend seen for miR172 is slightly different in this analysis than it was in the qPCR results. Here it appears that miR172 is slightly higher in the transgenic on July 3rd, the control on July 13th, and then is expressed equally between the two on the remaining sampling dates (Figure 1.10, Figure 1.11).

1.5 Discussion

AP2 genes link metabolism and development and the overexpression of *gl15* causes a delay in flowering. Pre-flowering, both the transgenic and the control accumulate sugar in their stems mainly in the forms of glucose and fructose; however, after flowering there is a shift towards production of sucrose. The shift is consistently seen in the composition and RNAseq analysis. This could mean that gl15 directly regulates all these genes or it could change a part of the pathway higher up and causes a cascading effect downward.

There could be two explanations for the 14 genes that are differentially expressed at all sampling dates (Table 1.1). One is that these genes are regulated strongly by *gl15*, most of them would appear to have some type of regulatory function themselves. It seems unlikely though that so many of these genes just happen to be on chromosomes 2 and 8 so the second explanation is that these genes are in regions where the allele from the original H99 transformation line has different expression compared to B73, so this now shows up as differential expression in transgene versus no transgene lines. The best way to distinguish these possibilities is to test how these genes behave in *gl15* mutant compared to control.

The flowering delay is also likely correlated to the repression of *FT-like genes* in the transgenics. *FT* has been shown to be a key protein in the transition from vegetative to reproductive phase and *CO* positively regulates *FT* in *Arabidopsis* (Kardailsky et al., 1999; Kobayashi et al., 1999; Samach et al., 2000; Wigge et al., 2005). *ZCN8* is believed to be the maize floral activator gene similar to *FT* in

Arabidopsis while *ZCN15* was ruled out due its predominant expression in kernels and not leaves (Danilevskaya et al., 2008; Meng et al., 2011). However, *ZCN15* is also playing a role in stems as well in our data. It would appear that the control maize operates in a similar fashion to *Arabidopsis* for the control of *FT* by *CO*. At flowering, levels of *CO-like* increase by almost eight fold and this is followed by an increase in *ZCN8* and *ZCN15*. Glucose may be acting as a signal for flowering in this case. The control accumulates glucose and once it reaches a certain level flowering occurs and there is a shift to sucrose production and a drastic increase in *ZCN15* expression as glucose is depleted. In the transgenic, more glucose is needed to trigger the shift, so the glucose is not converted to sucrose as quickly and *ZCN15 and ZCN8* expression is repressed which correlates with a delay in flowering.

In the transgenic, *CO* does slightly increase at flowering but never achieves the expression levels of the control. Photoreceptors and the *GI* gene regulate *CO* (Valverde et al., 2004; Mizoguchi et al., 2005). *GI* is two fold lower in the transgenic on July 13th when the control is flowering but the transgenic plants do obtain approximately equal *GI* expression by August 9th after the they have flowered. It is not apparent what exactly is causing the depression of *GI* and *CO*, but *FT* could be directly targeted by *gl15*. AP2-like transcription factors such as APETALA 2 (AP2) itself, the three TARGET OF EAT (TOE) proteins (TOE1, TOE2, and TOE3), and SCHLAFMÜTZE (SMZ) and its paralog SCHNARCHZAPFEN (SNZ) have all been shown to be repressors of flowering (Park et al., 2002; Aukerman and Sakai, 2003; Schmid et al., 2003). These AP2 genes cause this repression in flowering by regulating *FT* expression (Mathieu et al., 2009). Since the other AP2 genes are expressed almost identically between the two lines, *gl15* would be a likely candidate for AP2 repression of *FT* (Figure 1.10).

The eventual downregulation of *gl15* that occurs after July 13th could be the result of a *CO* independent pathway to flowering that was discovered in *Arabidopsis*. In this pathway, *GI* causes an increase in *miR172* and *miR172* then binds to its AP2 targets downregulating them allowing *FT* to be expressed (Jung et al., 2007). This

could explain why the levels of miR172 are higher in the transgenic and spike up along with the increase in *GI-like* and *FT-like* on August 9th (Figure 1.10).

The delay in flowering also causes a delay in sugar conversion in the transgenic stems. Around flowering, the stems begin converting glucose and fructose into sucrose and eventually that sucrose is pumped up to the ears and converted into starch. The transgenic plants have reduced grain yield (Pulam, 2012) due to the flowering delay. They do not have a sink for their sugar on July 13th and the delayed transport of sugar to the ears results in stems with higher sugar levels. Higher soluble sugar in the stems will make it easier for the transgenic stover to be converted to ethanol, making it better biofuel feedstock.

1.6 REFERENCES

Allsopp, A. (1967) Heteroblastic Development in Vascular Plants. *Advances in Morphogenesis* **6**, 127-171.

Anders, S., Pyl, P.T., & Huber, W. (2014) HTSeq — A Python framework to work with high-throughput sequencing data. *bioRxiv* preprint, doi: 10.1101/002824

Aukerman, M.J. & Sakai, H. (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15**: 2730–2741.

Bauer, S. & Ibáñez, A.B. (2014) Rapid determination of cellulose. *Biotechnology and Bioengineering*, in press.

Bolger, A.M., Lohse, M., & Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* btu170.

Cardon, G.W., Hohmann, S., Nettesheim, K., Saedler, H., & Huijser, P. (1997) Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in floral transition. *Plant J*. **12**, 367-377.

Chang, S., Puryear, J., & Cairney, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**, 113–116

Chen, X. (2004) A microRNA as a translational repressor of APETALA 2 in Arabidopsis flower development. *Science* **303**, 2022-2025.

Chuck, G., Cigan, A.M., Saeteurn K., & Hake, S. (2007) The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. *Nature Genetics* **39**, 544-549.

Chuck, G.S., Tobias, C., Sun, L., Kraemer, F., Li, C., Dibble, D., ... & Hake, S. (2011) Overexpression of the maize Corngrass1 microRNA prevents flowering, improves digestibility, and increases starch content of switchgrass. *PNAS* **108**, 17550-17555.

Danilevskaya, O.N., Meng, X., Hou, Z., Ananiev, E.V., & Simmons, C.R. (2008) A genomic and expression compendium of the expanded PEBP gene family from maize. *Plant physiology* **146**, 250-264.

Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J. A., & Paz-Ares, J. (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet*. **39**, 1033-1037.

Gandikota, M., Birkenbihl, R.P., Hohmann, S., Cardon, G.H., Saedler, H., & Huijser, P. (2007) The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP gene SLP3 prevents early flowering by translational inhibition in seedlings. *Plant J.* **49**, 683-693.

Galili, T. (2014) dendextend: Extending R's dendrogram functionality. R package version 0.14.2. http://CRAN.R-project.org/package=dendextend

Goebel, K. (1900) Organography in Plants. Part I. General Organography (trans. Balfour, I. B.) (Clarendon, Oxford) (English)

Ibáñez, A.B. & Bauer, S. (2014) Downscaled method using glass microfiber filters for the determination of Klason lignin and structural carbohydrates. *Biomass and Bioenergy* **68**, 75-81

Jung, J.H., Seo, Y.H., Seo, P.J., Reyes, J.L., Yun, J., Chua, N.H., & Park, C.M. (2007) The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in *Arabidopsis*. *Plant Cell* **19**, 2736-2748.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., & Weigel, D. (1999) Activation Tagging of the Floral Inducer *FT*. *Science* **286**, 1962-1965.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36.

Kerstetter, R.A. & Poethig, R.S. (1998) The Specification of Leaf Identity During Shoot Development. *Annu. Rev. Cell Dev. Biol.* **14**, 373-398.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., & Araki, T. (1999) A Pair of Related Genes with Antagonistic Roles in Mediating Flowering Signals. *Science* **286**, 1960-1962. DOI: 10.1126/science.286.5446.1960.

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., & Yano, M. (2002) Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant and Cell Physiology* **43**, 1096-1105.

Komiya, R., Ikegami, A., Tamaki, S., Yokoi, S., & Shimamoto, K. (2008) Hd3a and RFT1 are essential for flowering in rice. *Development* **135**, 767-774.

Komiya, R., Yokoi, S., & Shimamoto, K. (2009) A gene network for long-day flowering activates RFT1 encoding a mobile flowering signal in rice. *Development* **136**, 3443-3450.

Kuchelmeister, C. & Bauer, S. (2014) Rapid small-scale determination of extractives in biomass. *Bioenergy Research,* DOI 10.1007/s12155-014-9493-x.

Langfelder, P. & Horvath., S. (2008) WGCNA: an R package for weighted correlation network analysis. BMC *Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559.

Langfelder, P. & Horvath, S. (2012) Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of Statistical Software* **46**, 1-17.

Lauter, N., Kampani, A., Carlson, S., Goebel, M., & Moose, S.P. (2005) microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. *PNAS* **102**, 9412-9417.

Lawson, E.J.R. & Poethig, R.S. (1995) Shoot development in plants: time for a change. *Trends Genet.* **11**, 263-268.

MacDonald, J.W. (2008). affycoretools: Functions useful for those doing repetitive analyses with Affymetrix GeneChips. Rpackage version 1.36.0.

Mathieu J., Yant L.J., Murdter F., Kuttner F., & Schmid M. (2009) Repression of flowering by the miR172 target SMZ. *PLoS Biol.* **7**: e1000148.

McCarthy, D.J., Chen, Y., & Smyth, GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**, 4288-4297.

Meng, X., Muszynski, M.G., & Danilevskaya, O.N. (2011) The FT-like ZCN8 gene functions as a floral activator and is involved in photoperiod sensitivity in maize. *The Plant Cell Online* **23**, 942-960.

Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J., & Coupland, G. (2005) Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in Arabidopsis. *Plant Cell* **17**, 2255–2270

Moose, S.P. & Sisco, P.H. (1994) *Glossy15* Controls the Epidermal Juvenile-to-Adult Phase Transition in Maize. *Plant Cell* **6**, 1343-1355.

Moose, S.P. & Sisco, P.H. (1996) *Glossy15*, an *APETALA2*-like gene from maize that regulates leaf epidermal cell identity. *Genes Dev.* **10**, 3018-3027.

Park, W., Li, J., Song, R., Messing, J., & Chen, X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr. Biol.* **12**, 1484–1495.

Ploner, A. (2014) *Heatplus: Heatmaps with row and/or column covariates and colored clusters*. R package version 2.12.0.

Poethig, R.S. (2003) Phase Change and the Regulation of Developmental Timing in Plants. *Science* **301**, 334-336.

Pulam, R. (2012) The effect of shoot maturation on total biomass production in maize. Master's Thesis. University of Illinois Urbana-Champaign: U.S.A. http://hdl.handle.net/2142/29429.

Robinson, M.D. & Smyth, G.K. (2007) Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881-2887.

Robinson, M.D. & Smyth, G.K. (2008) Small sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, 321-332.

Robinson, M.D., McCarthy, D.J., & Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.

Robinson, M.D. & Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**, R25.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., & Coupland, G. (2000) Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science* **288**, 1613–1616.

Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D., & Lohmann, J.U. (2003) Dissection of floral induction pathways using global expression analysis. *Development* **130**, 6001–6012.

Schnable, P.S., Ware. D., Fulton, R.S., Stein, J.C., Wei, F, et al. (2009) The B73 maize genome: complexity, diversity and dynamics. *Science* **326**: 1112–1115.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., & Weigel, D. (2005) Specific effects of microRNAs on the plant transcriptome. *Dev. Cell* **8**, 517-527.

Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D. (2012) Determination of Structural Carbohydrates and Lignin in Biomass. Laboratory Analytical Procedure (LAP). National Renewable Energy Laboratory (NREL), Golden, CO. Revised Version 2012. http://www.nrel.gov/docs/gen/fy13/42618.pdf. Accessed March 2014.

Smyth, G.K. (2005) Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor'. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds), Springer, New York, 397-420.

Todesco, M., Rubio-Somoza, I., Paz-Ares, J., & Weigel, D. (2010) A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana. *PLoS Genet*. **6**, e1001031.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., Coupland, G. (2004) Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* **303**, 1003–1006.

Wang, J.W., Czech, B., & Weigel, D. (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. *Cell* **138**, 738-749.

Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., & Weigel, D. (2005) Integration of spatial and temporal information during floral induction in Arabidopsis. *Science* **309**, 1056–1059.

Wu, G., & Poethig, R.S. (2006) Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. *Development* **133**, 3539-3547.

Wu, G., Park, M.Y., Conway, S.R., Wang, J.W., Weigel, D., & Poethig, R.S. (2009) The sequential action of miR156 and miR172 regulates development timing in Arabidopsis. *Cell* **138**, 750-759.

Yang, L., Conway, S.R., & Poethig, R.S. (2011) Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. *Development* **138**, 245-249.

Yang, L., Xu, M., Koo, Y., He, J., & Poethig, R.S. (2013) Sugar promotes vegetative phase change in Arabidopsis thaliana by repressing the expression of MIR156A and MIR156C. *Elife*, **2**.

CHAPTER II
COMPARATIVE ANALYSIS OF STEM CARBON PARTITIONING, GENE NETWORKS, AND DEVELOPMENT IN RELATED SWEET AND ENERGY TYPE *SACCHARUM* LINES

2.1 ABSTRACT

Saccharum sucrose and fiber types are of high interest for their use in the biofuels industry. Although there is a long history of humans breeding *Saccharum* for high sucrose content, little is known about the mechanism that causes the sucrose to accumulate in the stems. Transgenic maize has shown that AP2 genes can delay flowering and increase stem sugar. The higher expression of AP2 genes in sugarcane bottom internode tissue could be a possible cause of sugar accumulation. Furthermore, the sugar accumulation could be directly upregulating *AINTEGUMENTA*-like (ANT), expansins, and *XYLOGLUCAN ENDO-TRANSGLYCOSYLASE/HYDROLASE* (*XTH*) to increase the size of the sink, stem internodes.

2.2 INTRODUCTION

The majority of carbon fixed in higher plants is used as material for synthesis of cell walls (Carpita & Gibeaut, 1993; Reiter, 1998). The main function of cell wall is structural support of the plant. Cell walls are a complex mix of polysaccharides and proteins. The main polysaccharide in plant cell walls is cellulose. Cellulose accounts for about one-third of the total mass of many plants. Hydrogen-bonded, linear chains of beta-linked glucose molecules form polymers where every other glucose molecule is rotated approximately 180 degrees to form a flat, smooth ribbon. These ribbons are called cellulose microfibrils and wrap around cells in overlapping layers providing resistance to osmotic pressure. This resistance allows plants to adopt an erect growth habit. Hemicellulose is a branched glucose polymer that is shorter than cellulose. They bind with pectin to cellulose to form cross-linked fibers. Lignins are covalently bonded to hemicellulose and add structural support to the plant. They fill in the spaces of the

cell wall between cellulose and hemicellulose. To keep up with the world's demand for energy sources, new methods are being explored. One of these methods is converting plant material, mainly cellulose and hemicellulose, into fuel.

A wide variety of plant species have been proposed as potential feedstock crops for production of biofuels from lignocellulosic biomass. Among these, four closely related species within the Andropogoneae tribe emerge as top candidates because of their superior efficiencies in photosynthesis, nitrogen economy, and water use efficiency. Maize, *Saccharum*, and sorghum are established and highly productive biomass crops. According to the Food and Agriculture Organization of the United Nations (FAO), *Saccharum* was the world's largest crop producing 1.8 billion tons on 25 million hectares in 2011 (http://faostat3.fao.org). Collectively, they are cultivated across the global spectrum of agriculture production environments. *Miscanthus* also shows promise as a leading feedstock crop based on its exceptional biomass yields with minimal production inputs in recent European and US field trials (Scurlock, 1998; Schwarz et al., 1998). Each of these closely related grasses also offer complementary advantages to diverse climates. Maize has proven successful in both tropical and temperate regions. Sorghum is known to be drought tolerant and varieties of *Miscanthus* have chilling tolerance. Maize and sorghum also have fully sequenced genomes (Schnable et al., 2009; Paterson et al., 2009). Some genomic sequences and transcriptome data is also available for *Saccharum* (de Setta et al., 2014; Grativol et al., 2014; Cardoso-Silva et al., 2014). Another advantage of these grasses is that they offer a wide variety of harvestable carbon sources. Maize provides carbon in the form of food, fuel and feed. The grain types pump sucrose into the ears where starch, lipids and proteins are synthesized, and lignin and cellulose accumulate in the stalk. Field corn fills its kernels with starch but there is a block in the starch conversion pathway in sweet corn that causes the kernels to accumulate simple sugars instead. There are also types of Energy Maize that divert carbon from the ears and produce more biomass (White et al., 2012). Sorghum also has multiple routes for carbon storage. Similarly to maize, grain sorghum pumps sugars into the head, which is used by developing seeds to make starch and protein. There are also varieties of sweet sorghum that, much like

Saccharum, accumulate sugar in the their stems. *Miscanthus* is not an established crop like the other three so we are just beginning to see its potential. *Miscanthus* x *giganteus* is known to produce large amounts of above ground biomass but it also translocates significant amounts of carbon into the underground rhizome for storage.

The endosperm tissue functions to store carbon in the form of starch for seedling development. In maize, a mutant, *sugary1* (*su1*), was discovered that has a non-functioning starch debranching enzyme. This results in sucrose accumulation in the kernel that does not get converted to starch (Correns, 1901; Pan & Nelson, 1984).

Saccharum has been recognized as one of the most efficient crops in converting solar energy into chemical energy. Modern cultivars are hybrids derived from interspecific crosses between *S. officinarum* and *S. spontaneum* and result in a high degree of polyploidy and frequent aneuploidy. *S. officinarum* is the high sucrose content species and *S. spontaneum* provides stress and disease tolerance and high fiber for biomass. In an attempt to increase the genetic diversity and disease resistance in these hybrids, modern hybrids were crossed again with *S. spontaneum*. These crosses generated some progeny with lower sucrose content and higher fiber content that were undesirable to breeders looking for high sucrose producing plants. However, these high fiber hybrids, dubbed energy cane, have been gaining attention lately as potential cellulosic biofuel feedstock in tropical and subtropical regions. The USDA sugarcane breeding program has defined two types of energy cane. Type 1 varieties result from commercial sugarcane crosses that yield progeny with fiber content too high for commercial sugar production. Type 2 varieties are F1 or Backcross 1 hybrids that result from crossing commercial sugarcane with *S. spontaneum*, or with the related species *Miscanthus* or *Erianthus* (Richard, 2008).

Sugarcane stores sucrose in mature stem parenchyma cells. Higher levels of stored sucrose have been shown to be associated with lignification of the xylem of vascular bundles and the sclerenchymatic bundle sheath extending out into the storage parenchyma. The suberization of these cells begins in the paranchyma cells adjuacent

to the vascular bundle sheaths and spreads to the storage parenchyma with the majority of storage parenchyma being lignified in mature internodes (Jacobsen et al., 1992).

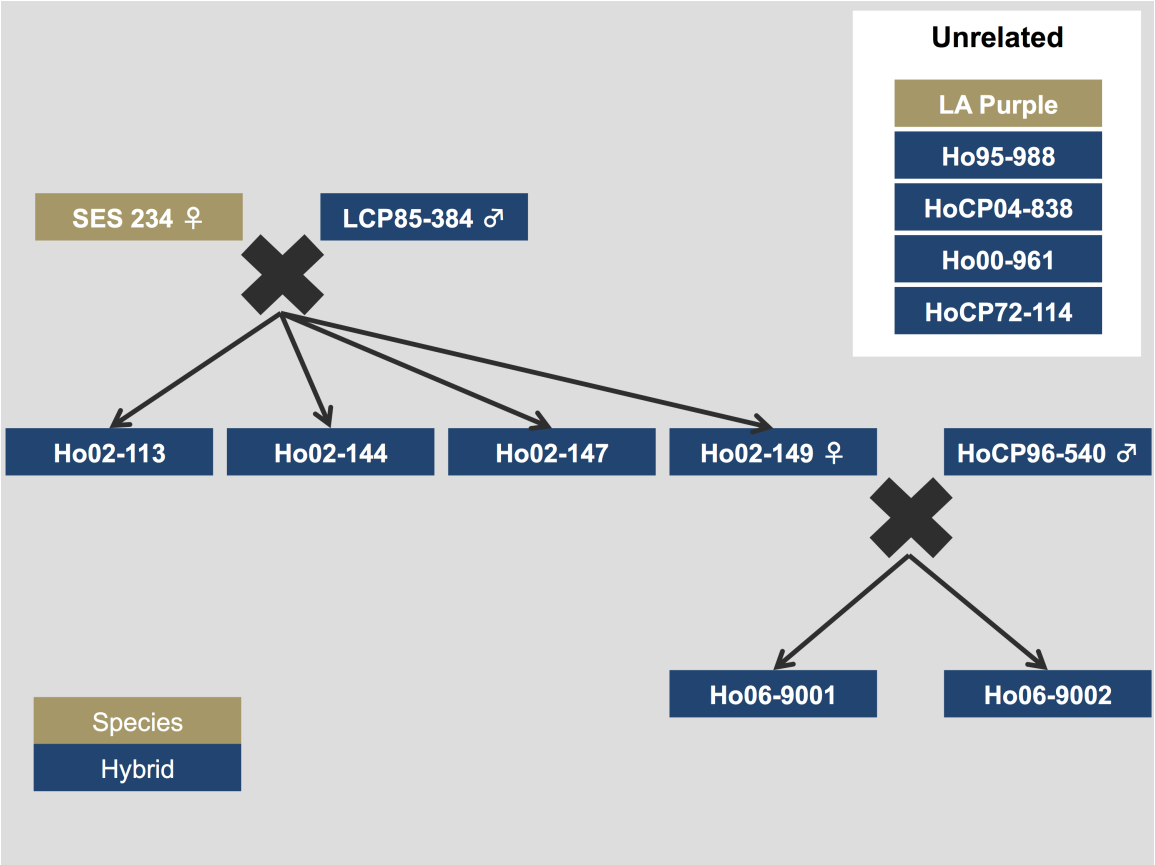
This work aims to understand the mechanisms that allow these grasses to store carbon with such versatility. The questions to ask are: is it a metabolic difference, a developmental timing variation, or a combination of the two.

2.3 MATERIALS AND METHODS

2.3.1 Plant Material

The USDA sugarcane basic breeding program in Houma provided 11 *Saccharum* lines: LCP 85-384, Ho 02-113, Ho 02-147, Ho 02-144, HoCP96-540, Ho 06-9001, Ho 06-9002, Ho 95-988, HoCP 04-838, Ho 00-961, and HoCP 72-114; the first seven listed of which are directly related. An additional twelfth line, *Saccharum officinarum* 'LA Purple', was also grown (Figure 2.1). The stem cuttings provided were cut so that a seed piece contained one node and an inch of internode on either side of the node. On November 25th of 2011, six seed pieces were planted per pot in a soil mix of 1:2:2 – Soil: Peat: Perlite. Four replicates of each line were grown in Urbana, Illinois at the University of Illinois Turner Hall Greenhouse.

Figure 2.1: *Saccharum* Lines. The blue boxes indicate *Saccharum* hybrids with varying fiber/sugar ratios. The brown boxes are considered pure species with SES 234 being an *S. spontaneum* and LA Purple being an *S. officinarum*. The box to the side shows unrelated lines and the pedigree shows the relatedness of the other varieties. Both female parents in the pedigree were not included in our experiment.



2.3.2 Anatomy Sections

On April 26th 2012, sections no more than one inch were cut from the first internode from the bottom, the tenth internode from the bottom, and the top 1 cm of LCP 85-384 and Ho 02-113 henceforth referred to as SC384 and EC113 respectively. These samples were locked into macro-cassettes and placed in 4% paraformaldehyde using the same method as described above in section 1.3.2. Three replicates were taken of every tissue type.

The sample collection bottle was placed uncapped in a desiccator for 24 hours under a vacuum. Samples were then dehydrated and cleared using a series of ethanol and xylene. Five grades of alcohol were used: 60%, 70%, 80%, 95% and absolute alcohol; and four grades of absolute alcohol and xylol were used: 25%, 50%, 75%, 100%. Each soaking stage lasted 24 hours and was performed in the desiccator under a vacuum. After clearing, paraffin wax chips (Fischer Scientific catalog # 22-050-128) were added to the xylene until the wax began melting. The samples were then transferred to a 58°C incubator and the wax was allowed to infiltrate for 24 hours. After this, the wax and xylene solution was poured off and replaced with fresh molten wax and allowed to infiltrate another 24 hours. Two more paraffin changes were performed (Langdon, 1920).

The *Saccharum* samples were then embedded, stained with Calcofluor White to look at cellulose and Schiff's reagent to look at lignin, and imaged using a confocal microscope. The methods for embedding, staining and imaging are described in section 1.2.3.

2.3.3 Composition Analysis

Top, middle, and bottom composition samples were collected from three replicate pots for each of the 11 *Saccharum* lines. Two dominant stems were selected from each pot and pooled. Top was defined as the top 10 cm, middle is the tenth internode from the bottom and bottom is the very first internode emerging from the soil. Samples were diced and placed in 50 ml centrifuge tubes. If there was excess sample tissue in the bottom and middle sections, equal amounts of each of the two pooled replicates were cut from the center of the sections and used for analysis. Sample analysis was performed using the methods described in section 1.3.3.

2.3.4 RNA Extraction and Sequencing

Samples collected for sequencing were taken from four replicate pots of SC384 and EC113. The tallest cane in each pot was sampled. The three tissue types collected were the top 1 cm of the plant, the tenth internode from the bottom, and the very bottom internode that had emerged from the soil (Figure 2.2). Samples were ground and RNA was isolated using the methods described in section 2.3.4.

Libraries for mRNA and small RNA were prepared and sequenced by the W.M. Keck Center at the University of Illinois. The methods were described in section 2.3.4; however, the mRNA libraries were paired end sequenced, only six samples were sequenced in each lane, and each sample was sequenced in two lanes to account for the complexity of the sugarcane genome.

Figure 2.2: Source Material for Sequencing. Top, middle, and bottom samples were cut from EC113 and SC384 for both mRNA and small RNA sequencing.

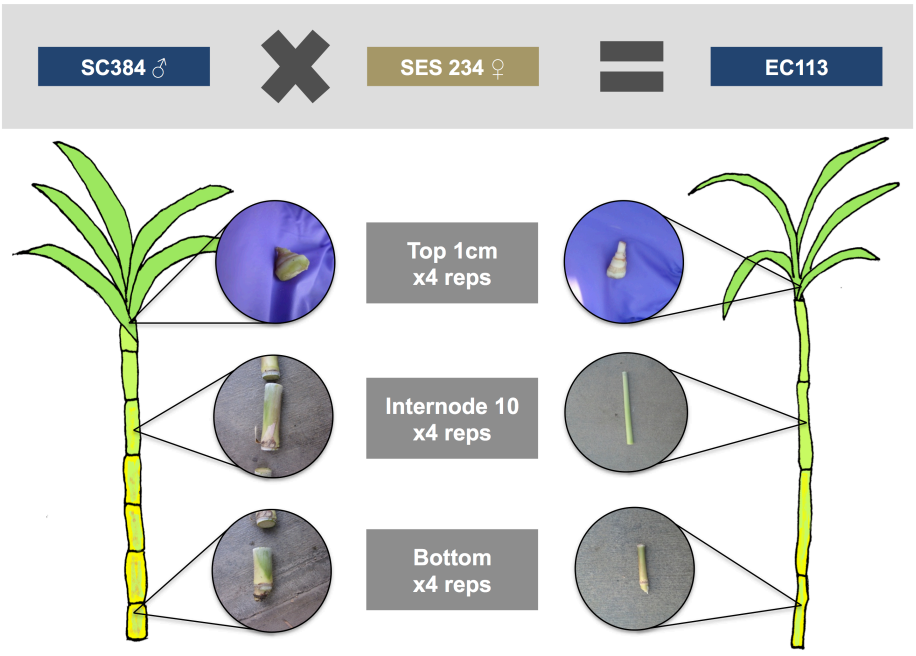
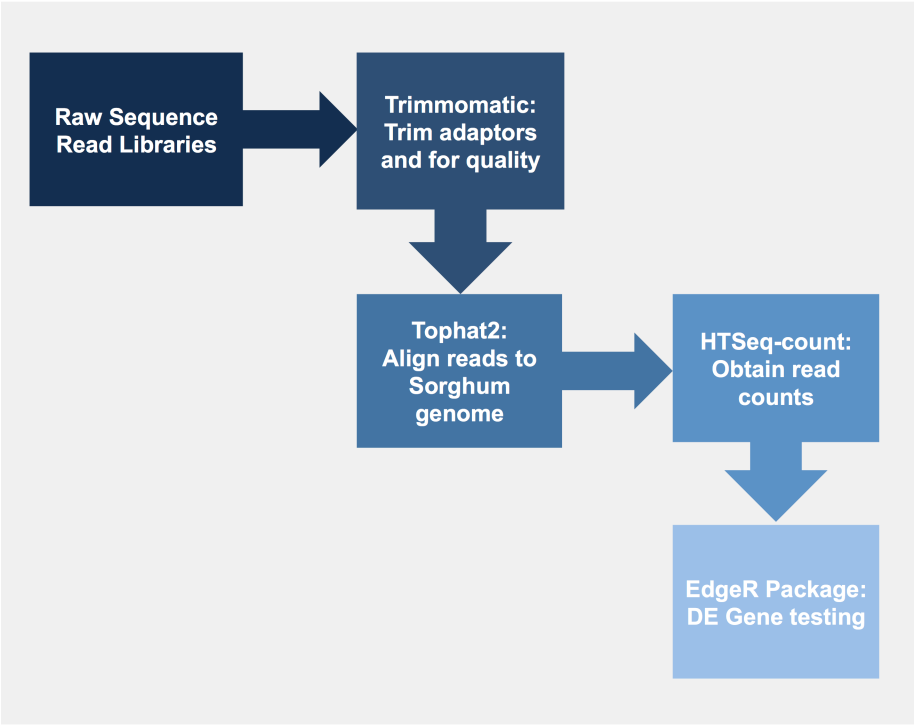


Figure 2.3: Alignment and Differential Gene Testing Pipeline. Raw reads were trimmed using Trimmomatic-0.30 and then aligned using TopHat v2.0.8. HTSeq was then used to count gene features and only those reads aligning uniquely were counted. Differential gene expression was determined using the edgeR package.



2.3.5 mRNA-Seq Analysis

The pipeline for mRNA sequence analysis is shown in Figure 2.3. The 24 mRNA libraries generated almost 1.5 billion read pairs. The paired end mode of Trimmomatic-0.30 (Bolger, et al. 2014) was used to trim the adaptors and check for read quality. Unique adaptor files were created for each library to assist with trimming. Bases were removed if their phred score fell below 30 and the entire read was discarded if its length was less than 40 nt. After trimming, 98.9% of the total reads remained.

The reads were aligned to two references using TopHat v2.0.8 (Kim et al., 2013). The first reference used was the *Sorghum bicolor* v1.4 genome (ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Sbicolor_v1.4/; Paterson et al. 2009).

Sequences not assembled into chromosomes were removed for aligning. The following parameters were used: read-mismatches 10, read-gap-length 6, read-edit-dist 10, read-realign-edit-dist 0, min-intron-length 25, max-insertion-length 15, mate-inner-dist 200, mate-std-dev 100, max-deletion-length 15, max-multihits 10, microexon-search, b2-very-sensitive. Tophat threw out about 1% of the trimmed reads because they were missing their mate. The remaining reads mapped to sorghum at a rate of 73.3% with 91.4% of those mapping uniquely and 83.5% mapping as concordant pairs.

The second reference used was the third version assembly from a doubled haploid *Miscanthus sinensis* (EBI unpublished data). It was aligned using the same Tophat parameters as the sorghum alignments. The mapping rate was 82.3% with a unique mapping rate of 84.2% and a 68.0% concordant pairs mapping rate.

The sorghum aligned reads were counted using HTSeq (Anders, et al. 2014). HTSeq counted uniquely mapped reads with a mapping quality score of 50 at the gene level using the intersection_nonempty mode with gene coordinates based on the *Sorghum bicolor* v1.4 annotation. The resulting libraries averaged 64% of the mapped reads.

The same six Bioconductor packages that were described in section 1.3.5 were used to examine reads and determine differential expression for the sorghum alignments. 27,412 chromosomal sorghum gene models were considered and of these, 25,589 had at least one read that mapped in at least one sample. A gene model was considered for differential expression analysis if it showed an average CPM of five or higher in at least three samples. The 5 CPM threshold gave the data a normal distribution and left 16,200 gene models and at least 98% of the reads in all libraries (Figure 2.4). The PCA and hierarchical cluster analysis showed that middle samples were in a state of flux with some behaving more like tops, some being an intermediate state, and others behaving more like bottoms. For this reason, they were discarded. One entire stem was also removed from the remaining top and bottom samples to reduce variation in the samples and tighten up the PCA plot. This left three top and

bottom samples for each *Saccharum* variety for differential expression analysis (Figure 2.5). The data show that the varieties are very similar and that the majority of the difference is within stems with the first principal component being internode age and the second principal component being sugar/fiber content.

The contrast matrix for this experiment made four comparisons: the difference in top samples between varieties (Top_SC384/EC113), the difference in bottom samples between varieties (Bottom_SC384/EC113), top versus bottom within SC384 (SC384_Top/Bot), and top versus bottom within EC113 (EC113_Top/Bot). The edgeR package was again used to further normalize the data and determine p-values. To be considered differentially expressed, a gene needed an FDR less than or equal to 0.001. Fold change and minimum CPM were determined based on MA plots with the CPM cutoff corresponding to the value of the upregulated sample (Figure 2.6). The minimum fold change and CPM values were as follows: Tops, 2 and 10; Bottoms, 3 and 15; LCP_TvB, 4 and 20; and Ho02_TvB, 4 and 20.

Figure 2.4: Counts per million gene filtering of reads. A 5 CPM threshold was chosen because it gave the kept genes an approximately normal distribution then can be seen when comparing the difference between the distribution before and after filtering shown in panels A and B respectively. Panel C shows genes that were discarded.

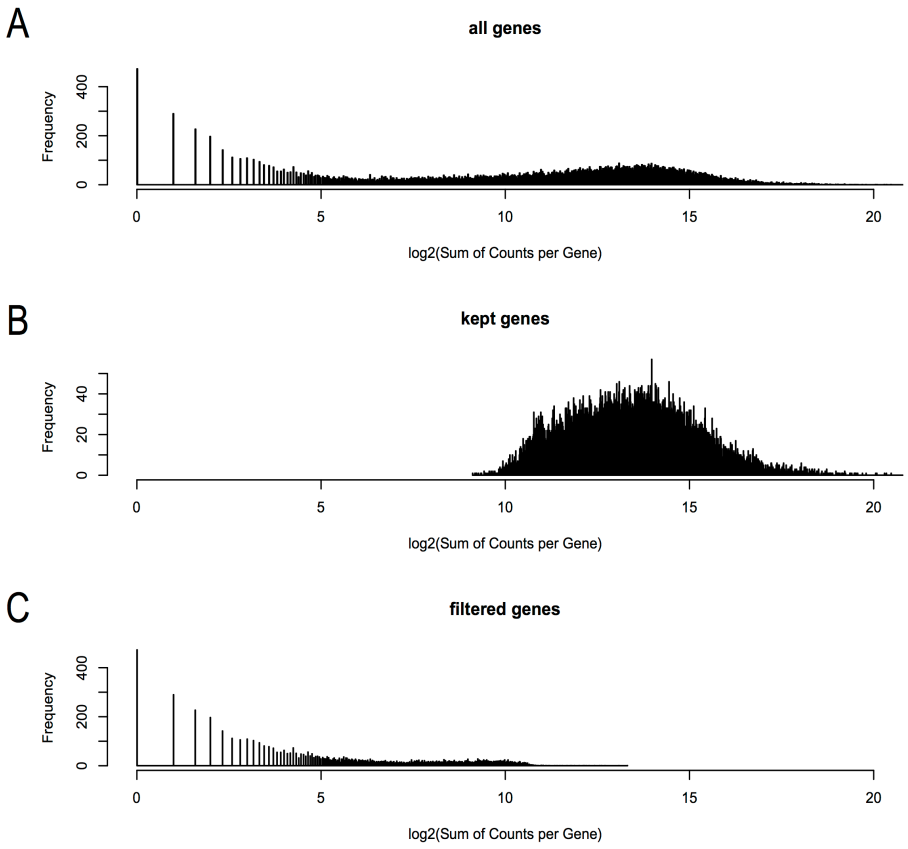


Figure 2.5: edgeR Normalized PCA and Hierarchical Clustering Tree for the sorghum aligned libraries. Panel A shows the principal component analysis for all of the sequenced libraries. The middle tissue samples show a wide range of distribution. Once the middle tissue samples and one outlier from each top and bottom tissue sample were removed, the improved principal component plot shows that the first and second sources of variation in the experiment come from internode age and sugar/fiber content respectively. Panel C shows confirms this result showing that the main branches of the hierarchical cluster are internode age and the sub branches are sugar/fiber content.

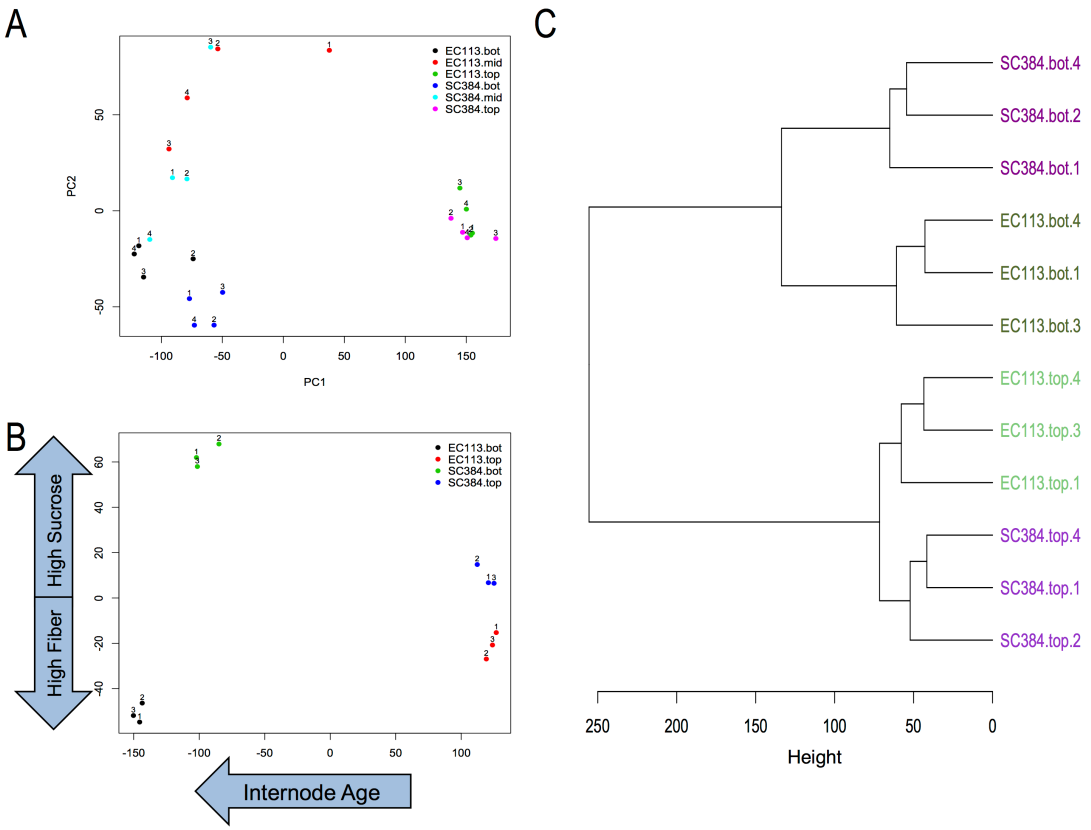
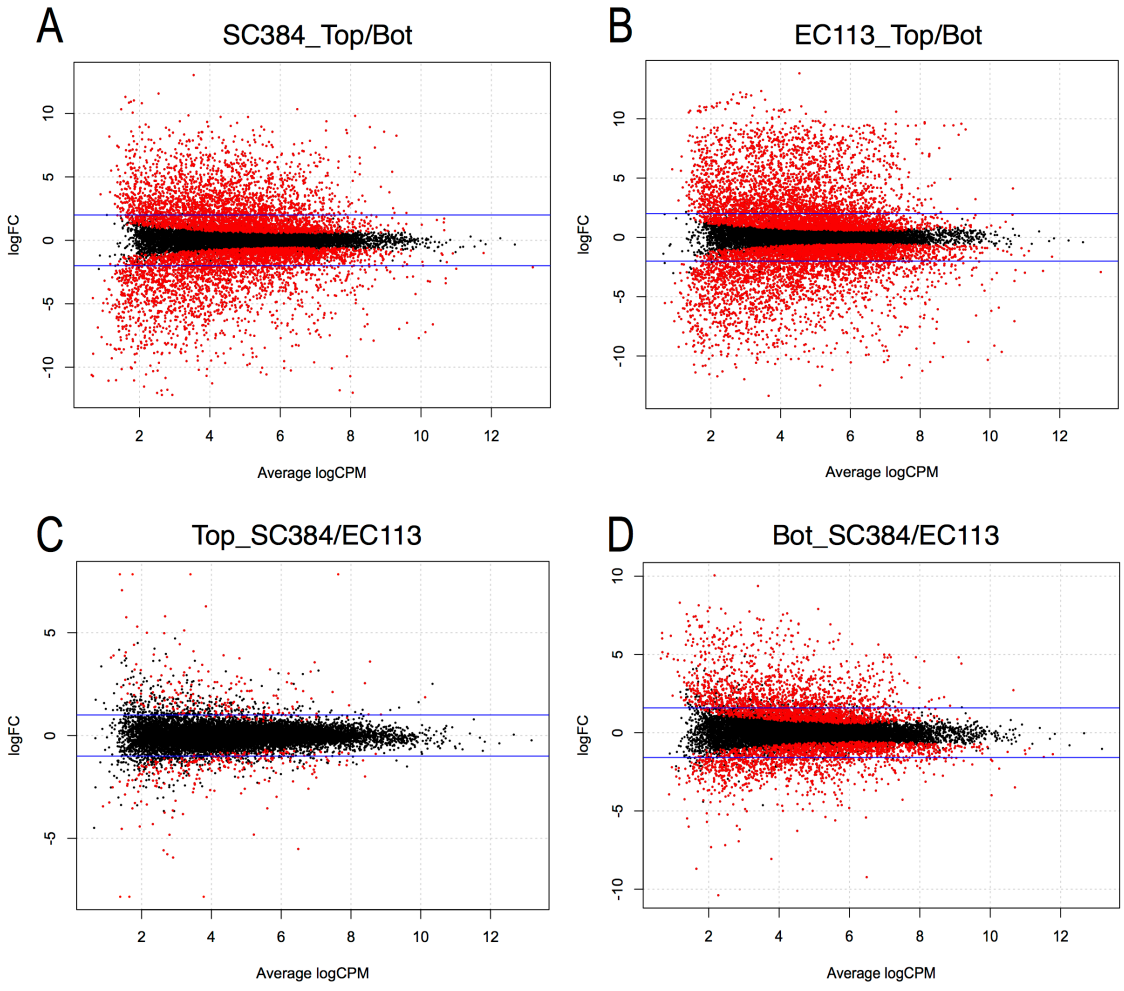


Figure 2.6: MA plots. The red dots indicate genes that have a FDR p-value less than or equal to 0.001. The blue lines indicate the logFC (log fold change) cutoff that was selected for each group. A logFC of 4 was selected for panel A and B and logFC 2 and 3 were selected for panels C and D respectively due to the differences in the patterns of dispersion. Panels A and B show the within variety differences between top and bottom tissues for SC384 and EC113. The tissue specific differences between varieties are show in panels C and D. The variance among genes for the comparisons within a variety is greater than comparing the same source tissue between the varieties.



2.3.6 miRNA-Seq

The same pipeline was used for aligning and counting the small RNA data that was described in section 1.3.6. The 24 libraries generated 681 million reads. They

were sequenced in the same lanes as the maize small RNA. Each miRNA in every sample had the counts per million value calculated and then transformed by taking the logarithm base 2. These values were then imported in R and a heatmap was constructed using the Heatplus package (Ploner 2014).

2.3.7 qPCR Analysis

Four potential housekeeping genes and 13 genes of interest were chosen from the sorghum alignments for further qPCR validation. Gene coordinates were obtained for each gene from the sorghum general feature format file (GFF) and a Perl script was written to extract from the SAM file all the read sequences aligned to these positions. The reads were then aligned again to the coding sequence using Sequencher 5.0.1. A consensus strained was used to design the primer pairs and the following criteria were entered to using the IDT PrimerQuest website (http://www.idtdna.com/Primerquest/Home/Index): fragment length 80 to 200 nt, primer length 17 to 28 nt, Tm 63° to 66°C, and GC content 41-65%. The primers designed by the program where then aligned to original sequence. Pairs were only considered if they were located on the 3 prime end of the sequence, did not overlap a conserved protein domain, and shared 100% identity with all reads and the sorghum reference gene. Novoalign v3.00.02 was used to align the primer pairs to the *Sorghum bicolor* v1.4 genome. All alignments were reported and the -l and –i were set based on the size of the primers and the expected genomic size of the PCR fragment. Any primer pair that had more than one mapping location or did not map to its expected gene location was discarded. The pairs were then aligned to the *Miscanthus sinensis* v3 double haploid genomic assembly. It was assumed that *Miscanthus* and *Saccharum* share the same duplication event so that each primer pair should map twice to the assembly to account for paralogs so any primer pair mapping more or less than twice was discarded. After filtering, 25 primer pairs were identified for further testing (Table 2.1).

RNA expression values of the genes of interest were quantified using the Lightcycler® 480 System (Roche). Top and bottom tissue that was sampled along with

and in the same manner as the RNA sequencing tissue was collected from LCP 85-384, Ho 02-113, HoCP96-540, and Ho 06-9001. Three biological replicates were sampled for each cane. RNA samples were reverse transcribed to cDNA using the Invitrogen® SuperScript® III First-Strand. The following 20 µL reaction mixture was used: 10 µL LightCycler® 480 SYBR Green I Master (Roche), 7 µL H2O, 2 µL 1:6 dilution of cDNA, 1 µL 10 µM forward and reverse primers (Table 2.1). Reaction mixes were subjected to 45 cycles of cDNA amplication, followed by a standard melting curve. The primers amplifying glyceraldehyde 3-phosphate dehydrogenase were used as the relative expression control in the study as they had the most even expression in all samples. Cycle threshold (C_T) for each sample was calculated using the method described in section 1.3.7.

The three biological replicates from each variety used in the above mRNA qpcr experiment were quantified for the expression of miRNA 156, miRNA 172, and miRNA 168 using the techniques described in 1.3.7. The same kits, primers, and other reagents were used and analysis was done with the same equipment and Perl Script.

Table 2.1: qPCR Primers. Pair pairs designed using *Saccharum*, *Sorghum* and *Miscanthus* sequences. Primer pairs were designed from *Saccharum* mRNA sequence reads. Reads were aligned to the *Sorghum* genome and the reference was then used to design primers. Pair sequences must align to sorghum only once and the DH1 *Miscanthus* scaffolds twice, to account for genome duplication.

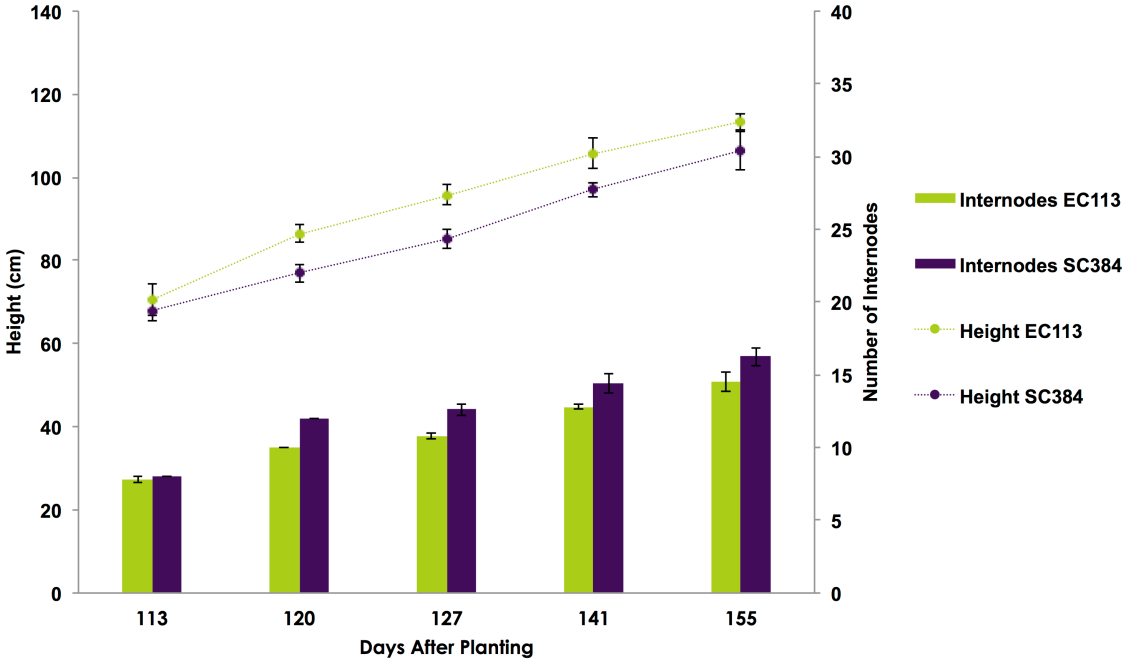
Name	Sorghum ID	Fragment Size	Length	Tm	GC%	Sequence	DH1v2 Scaffold Matches
ADF-1F	Sb07g012320	134	26	65	42.3	GGGTACTTCATCAAAGGTGGAATTGT	Scaffold468099,Scaffold366851
ADF-1R	Sb07g012320	134	23	65	47.8	CAAGCATGTAACCAGGTGCTGAT	Scaffold468099,Scaffold366851
ADF-2F	Sb07g012320	133	28	65	42.9	GCTCGAATTGGAGACAATGTGAAGATAC	Scaffold468099,Scaffold366851
ADF-2R	Sb07g012320	133	26	65	50	CTGTTCCACTAGGGAGTAAAGCATCC	Scaffold468099,Scaffold366851
ADF-3F	Sb07g012320	102	27	65	40.7	TGCTTTACTCCCTAGTGGAACAGTTAT	Scaffold468099,Scaffold366851
ADF-3R	Sb07g012320	102	24	65	45.8	TCTGCAGATCAAGCATGTAACCAG	Scaffold468099,Scaffold366851
CDF-1F	Sb02g043310	155	21	66	52.4	ACCCTGTCGAAGCTGCAAAGA	Scaffold66550,Scaffold44121
CDF-1R	Sb02g043310	155	22	66	50	TGACAGAGCTGCTGGATTTGCT	Scaffold66550,Scaffold44121
CDF-2F	Sb02g043310	110	23	66	43.5	TGTTCAATGGAATGTGCCACCAA	Scaffold66550,Scaffold44121
CDF-2R	Sb02g043310	110	24	66	45.8	GCCAAGGAGTAATGAATGGCCAAA	Scaffold66550,Scaffold44121
CESA-F	Sb04g006260	194	20	65	55	TGGAAGTCGGCGTCGAAGAT	Scaffold203857,Scaffold3969
CESA-R	Sb04g006260	194	21	65	57.1	GACGACTCCACGGACGAAATG	Scaffold203857,Scaffold3969
CO-1F	Sb04g025660	169	17	64	64.7	CCGAGCTGGACATCGCT	Scaffold184132,Scaffold363566
CO-1R	Sb04g025660	169	20	64	55	TTCTCACGGTACCGCATCAG	Scaffold184132,Scaffold363566
GLU-F	Sb07g026520	97	20	65	55	TTCACGCACGCCAATGTCAG	Scaffold366010,Scaffold21812
GLU-R	Sb07g026520	97	21	65	57.1	CGAACCAGTCGACGAAATGGC	Scaffold366010,Scaffold21812
IDS-1F	Sb01g003400	118	22	66	54.5	GTTGAAGCGGCAAGGGCTTATG	Scaffold144262,Scaffold311095
IDS-1R	Sb01g003400	118	22	66	50	TTTCGGTGTCAGGCAGGCTATT	Scaffold144262,Scaffold311095
IDS-2F	Sb01g003400	152	23	66	52.2	ATACGATGGCCTCTCAGCCAATG	Scaffold144262,Scaffold311095
IDS-2R	Sb01g003400	152	20	66	55	TCGCTCCATTGGCCTTTCCT	Scaffold144262,Scaffold311095
ANT6-1F	Sb06g031120	136	26	66	46.2	CCATCATAGCTTTAGTCAGCCCTGTA	Scaffold191975, Scaffold609799
ANT6-1R	Sb06g031120	136	25	66	48	GGGCCAAACTAAAGTGACCAGAAAG	Scaffold191975, Scaffold609799

Table 2.1 (cont.)

ANT6-2F	Sb06g031120	190	25	66	48	GCTTTAGTCAGCCCTGTACAATGGA	Scaffold191975,Scaffold609799
ANT6-2R	Sb06g031120	190	26	66	42.3	TGTTCAGTACAAGACAGGAGCTTTGA	Scaffold191975,Scaffold609799
ANT10-F	Sb10g026150	181	23	65	47.8	TACAGAGGAGTCACAAGGCATCA	Scaffold99154, Scaffold34244
ANT10-R	Sb10g026150	181	22	65	54.5	GGCTCATGTCGAAGTTGGTCAC	Scaffold99154, Scaffold34244
SPL-F	Sb07g027740	80	26	66	50	CTTCTGTCAACTCAGCCATGGGATAC	Scaffold583319,Scaffold109764
SPL-R	Sb07g027740	80	22	66	50	AAACCTGCTGTTGCAGGCATTG	Scaffold583319,Scaffold109764
Starch-F	Sb10g008200	137	20	63	55	CTCGACACGTACCGGAACTA	Scaffold208368, Scaffold574553
Starch-R	Sb10g008200	137	20	63	50	AGGGTTCACCACTGGTACTT	Scaffold208368, Scaffold574553
SUS-F	Sb01g033060	170	22	66	50	TGGCACATGAGATTGCTGGAGA	Scaffold405206,Scaffold423147
SUS-R	Sb01g033060	170	27	66	48.1	CCAGTAGAGGTCAGAGTTAGGGTACTT	Scaffold405206,Scaffold423147
UDP-F	Sb01g007580	167	24	66	50	CAGTGAGCGTACACATCCTTGAGA	Scaffold480665,Scaffold488038
UDP-R	Sb01g007580	167	24	66	50	CACAACAGCAAGGGCATCAACTAC	Scaffold480665,Scaffold488038
VGT-F	Sb09g002080	136	24	65	50	AAGAAGAACCAGAGGAGGAAGAGG	Scaffold379916, Scaffold445254
VGT-R	Sb09g002080	136	18	65	61.1	GGTTGCGTGCAGGATGGT	Scaffold379916, Scaffold445254
GAPDH-F	Sb04g025122	95	23	64	47.8	ACTGTTAGGCTTGAGAAGTCTGC	Scaffold36368,Scaffold325194
GAPDH-R	Sb04g025122	95	23	65	47.8	ACATAACCCAGAATGCCCTTGAG	Scaffold36368,Scaffold325194
E1-F	Sb05g00520	154	26	66	46.2	TCTCTGAGAGGAAGAAGTTGCTTGAC	Scaffold454959,Scaffold104446
E1-R	Sb05g00520	154	20	66	60	CTCTTCGGCTCCAACGTCCT	Scaffold454959,Scaffold104446
Tublin-1F	Sb01g006310	119	26	66	46.2	GCATTGTAAGGCTCCACAACAGTATC	Scaffold336669,Scaffold120665
Tublin-1R	Sb01g006310	119	22	66	50	GGCATGGGCACCCTTCTTATTT	Scaffold336669,Scaffold120665
Tublin-2F	Sb01g006310	192	22	66	54.5	CCACCTCCTTGGTGCTCATCTT	Scaffold336669,Scaffold120665
Tublin-2R	Sb01g006310	192	22	66	50	TCCACTTCTTCATGGTCGGCTT	Scaffold336669,Scaffold120665
Tublin2-1F	Sb04g037170	135	22	66	50	ACTGCCACAAGCGACTTGAAAC	Scaffold26510,Scaffold225455
Tublin2-1R	Sb04g037170	135	24	66	50	GCTTGACAACTGCTTCAACCTCTG	Scaffold26510,Scaffold225455
Tublin2-2F	Sb04g037170	191	22	66	54.5	CGTGCAAGAGGATCGTGAAGGA	Scaffold26510,Scaffold225455
Tublin2-2R	Sb04g037170	191	24	66	50	CAAGTCAGCCAGTGCAGTTTCAAG	Scaffold26510,Scaffold225455
Tublin2-3F	Sb04g037170	151	21	65	57.1	CGCCGATCCCTACGATCTCAA	Scaffold26510,Scaffold225455
Tublin2-3R	Sb04g037170	151	24	65	50	CTCAGCACCTTGCTCATTTGACTC	Scaffold26510,Scaffold225455

2.4 RESULTS

Figure 2.7: Growth of sequenced Saccharum lines. EC113 and SC384 were measured for height and total internodes with emerged leaves over their growing period in the greenhouse. EC113 grew taller but SC384 produced more internodes. Error bars indicate standard error.



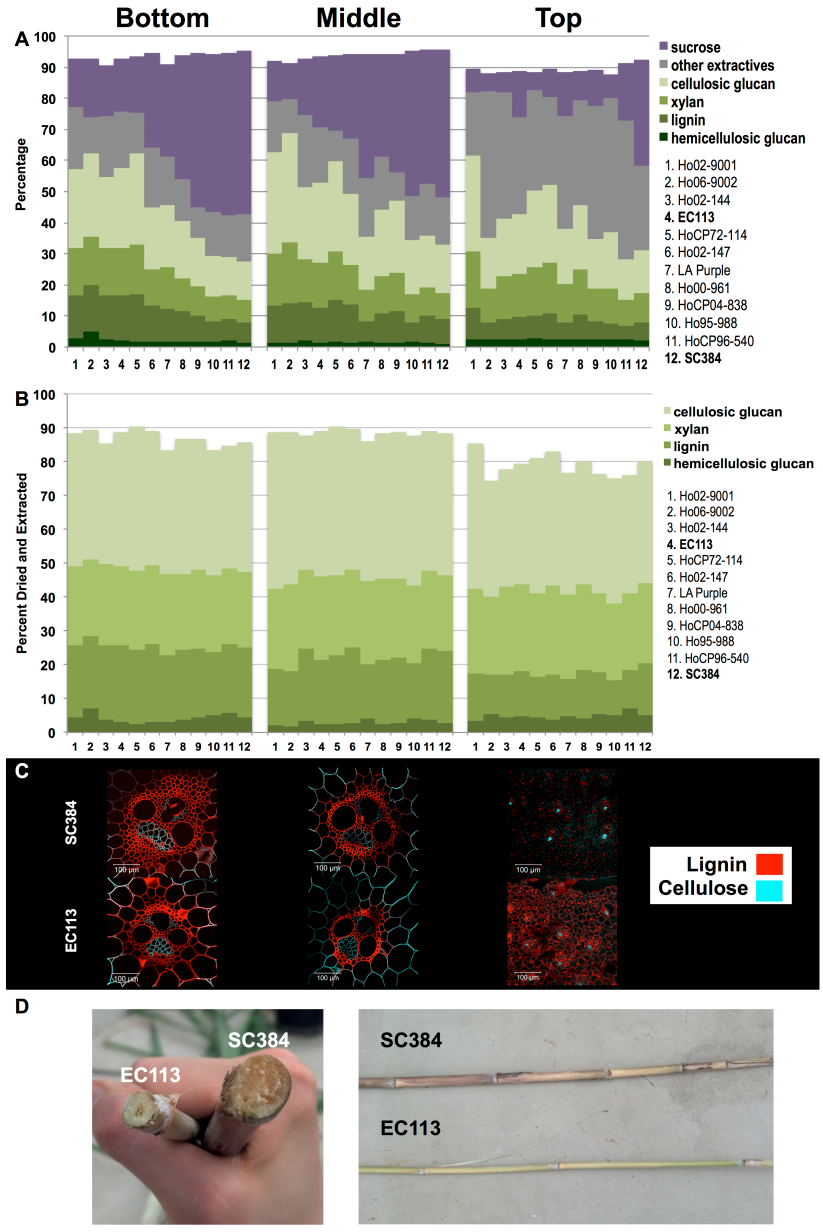
Periodically, the pots of EC113 and SC384 were examined in the greenhouse. Height was measured as the length from the soil level up to the leaf whorl. Internodes were counted as number of fully emerged leaves plus any additional visible internodes where leaves were missing. Only the largest cane in each pot was measured. Overall, EC113 produced taller stalks and SC384 produced more internodes (Figure 2.7). The EC113 canes stalks were much smaller in diameter and the plant produced more tillers and rhizomatous structures could be seen underneath the soil.

The composition analysis shows that there is a significant range of carbon partitioning among the *Saccharum* lines even between siblings. All varieties have more

sucrose in their bottoms and middles than in their tops. Most of the soluble sugar in the tops is other extracts like fructose and glucose. When looking at the bottom stems, five varieties: Ho 06-9001, Ho 06-9002, Ho 02-144, EC113 and HoCp 72-114, stand out for having the most fiber content due to their higher percentages of cellulosic glucan, xylan, and lignin. Commercial sugarcane varieties: HoCP 04-838, Ho 95-988, HoCP 96-540, and SCP384, have significantly more sucrose in their bottom stems. The *S. officinarum* "LA Purple" species type is intermediate between the high fiber types and the commercial sugarcanes when looking at the mature internode. If sugar content is removed and the plants are just looked at on a percent by dried and extracted scale, the amount of fiber portions of the stem: cellulosic glucan, xylan, lignan, and hemicellulosic glucan, become more even among all the lines (Figure 2.8).

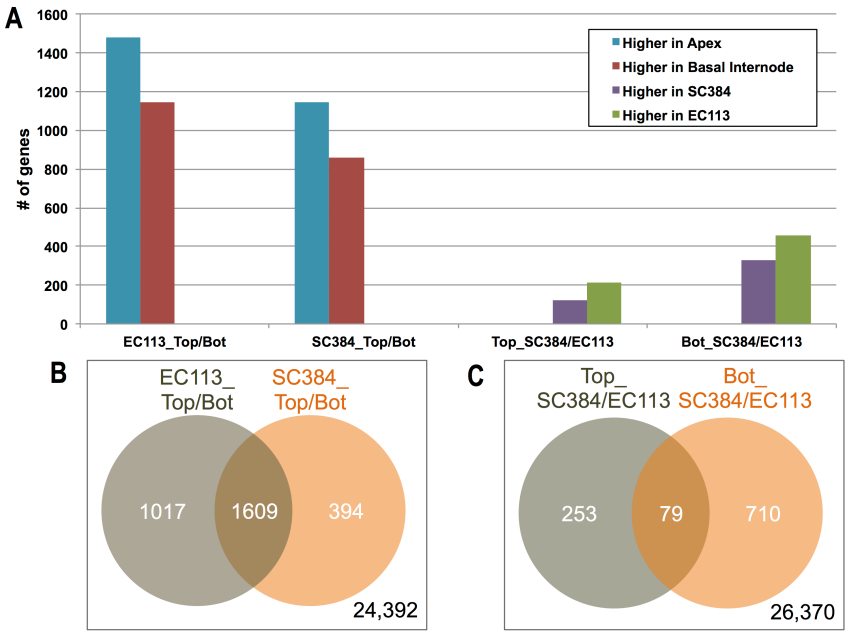
Overall, the diameter of the SC384 stems is thicker than that of EC113. The top sections stained for lignin and cellulose do appear to show that EC113 is slightly more lignified than SC384. Similar to the findings in Jacobsen et. al 1992, an increase in the lignification of cells surrounding the vascular tissue is observed as the stem tissue matures and sucrose content increases, with SC384, the higher sucrose storing variety, showing the most lignification (Figure 2.8).

Figure 2.8: Composition analysis of 12 *Saccharum* lines and anatomy sections of RNA sequenced varieties. Panel A shows the composition difference of carbon components that were quantified from the stems of our *Saccharum* population based on percent total. Percent by dried and extracted composition is showed in panel B. The results of the cellulose and lignin staining of top, middle, and bottom tissues of EC113 and SC384 is illustrated in panel C. Panel D is a side by side comparison of the thickness of EC113 and SC384 stems.



The mRNA sequencing differential gene analysis made four comparisons: within EC113 comparing top to bottom (EC113_Top/Bot), within SC384 comparing top to bottom (SC384_Top/Bot), comparing the top tissues of SC384 and EC113 (Top_SC384/EC113), and comparing the bottom tissues of SC384 to EC113 (Bot_SC384/EC113). The within stem comparisons, EC113_Top/Bot and SC384_Top/Bot, have slightly more genes expressed at higher levels in the top samples at 56% and 57% respectively with EC113_Top/Bot having 1.3 times more differentially expressed genes than SC384_Top/Bot. The overlap between the two within stem comparisons is 1,609 genes or 80% of SC384_Top/Bot and 61% of EC113_Top/Bot. When looking at Top_SC384/EC113 and Bot_SC384/EC113, the comparisons also show more genes with higher expression levels in EC113 with 64% and 58% respectively of the genes showing more expression in the progeny. However, there is a greater separation between the two comparisons with Bot_SC384/EC113 having 2.4 times more differentially expressed genes than Top_SC384/EC113. The overlap of the two comparisons only accounts for 24% in the genes in Top_SC384/EC113 and 10% in Bot_SC384/EC113 (Figure 2.9). In summary, the top tissues of both varieties have slightly larger amounts of genes with higher expression when compared to their bottom counterparts. EC113 shows more differences between its top and bottom tissues compared to SC384 which is seen both in the number of differentially expressed genes and in the overlap between the top comparisons. When looking at the tissue specific comparisons, there are fewer differences between tissues of different varieties than there are between the two tissue types. However, bottom tissues show more differences than tops.

Figure 2.9: Differential gene expression between the four mRNA sequencing comparisons. Panel A shows the number of genes differentially expressed in all comparisons. The comparisons are split into two bars with each bar representing the number of genes that were higher in that particular variable in the comparison. Panels B and C look at the overlap between the within stem comparisons and the between tissue comparisons respectively. There is much higher overlap between the within stem comparisons indicating that tissues of the same age from separate varieties are more similar than tissues from the same plant that are of different ages.

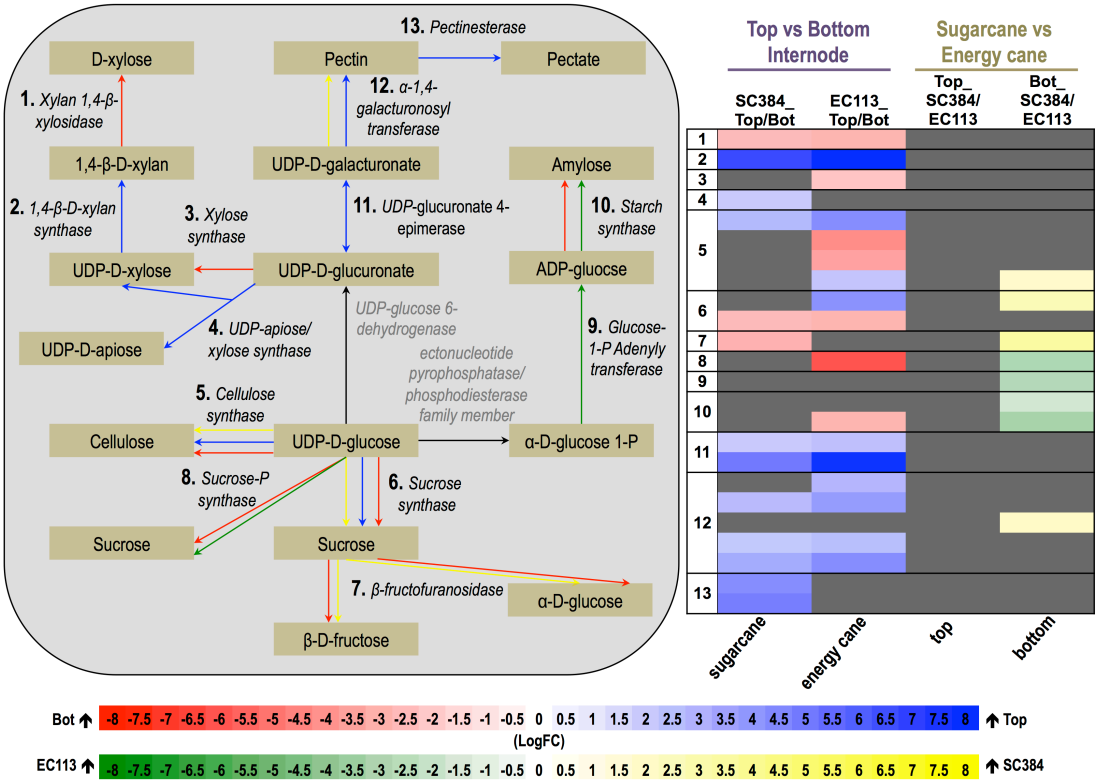


SC384 and EC113 differ in the way they store carbon as seen in the composition analysis. However, RNA-Seq analysis showed that no starch and sucrose metabolism genes were differentially expressed between the top tissues. In the bottom tissues, LCP showed genes with higher of cellulose synthase, cell wall invertase (β-fructofuranosidase), α-1,4-galacturonosyltransferase, and sucrose synthase. These enzymes are responsible for the breakdown and synthesis of sucrose, pectin, and cellulose. Sucrose-P synthase, glucose-1-P adenyly transferase, and starch synthase showed higher expressing in EC113 with the former playing a part in sucrose-

phosphate synthesis and the two later playing a part in converting glucose to amylose (Figure 2.10).

When comparing the genes differentially expressed in the starch and sucrose metabolism pathway for the Top_SC384/EC113 and Bot_SC384/EC113 comparisons, there are many similarities. One difference is that the bottoms of EC113 show higher expression of xylose synthase, sucrose-p synthase, and starch synthase genes while there was no difference in SC384 tissues for these genes. SC384 tops express higher levels of UDP-apiose/xylose synthase and pectinesterase and SC384 bottoms express more β-fructofuranosidase (Figure 2.10).

Figure 2.10: Starch and Sucrose Metabolism Pathway. The differential expression between the four comparisons is shown. Red arrows and boxes indicate that a gene was expressed higher in one or both top tissues compared to the bottom and blue boxes and arrows mean the opposite. Similarly, yellow boxes and arrows indicate that a gene has higher expression in the bottom tissue of SC384 compared to EC113 bottom and green indicates that expression was higher in EC113 bottoms. Gray boxes and arrows indicate no differential expression. All enzyme names and reaction directions are based on the chart at http://www.kegg.jp/kegg-bin/show_pathway?140807991418445/ko00500.args.



The RNA-Seq experiment showed that the bottom tissues of SC384 had higher expression of eight AP2 domain containing genes and three of these genes: Sb01g003400 similar to IDS1, Sb09g002080 related to AP2.7/TOE1, and Sb06g030670 APETALA 2-like; showed microRNA 172 (miR172) binding domains. Sb09g002080, related to AP2.7/TOE1, was expressed at higher levels in EC113 bottoms compared to

its top also. Of the other five AP2 domain containing genes, two of them, Sb10g026150 and Sb06g031120, are labeled as AINTEGUMENTA-like 5 and are both highly expressed in the bottom internodes of SC384 compared to all tissues with Sb10g026150 only being detected in that tissue. Six squamosa promoter-binding protein-like (SPL) transcription factor genes showed elevated expression levels in top tissues compared to their bottoms of both varieties but no difference was detected between varieties (Figure 2.11). SC384 bottoms and tops were expressing the cell wall targeting genes expansin and *XYLOGLUCAN ENDO-TRANSGLYCOSYLASE/HYDROLASE* (*XTH*) at higher levels than EC113 tops and bottoms but at equal or less levels than the top tissue of SC384 (Figure 2.12). Expression of Sb04g026320, ECERIFERUM 1-like (CER1) and WAX2-like, was considerably higher in top tissues compared to their bottoms; however, SC384 bottom had a nine fold increase in expression compared to EC113 bottom (Figure 2.13).

Figure 2.11: Expression of miRNA and SPL and AP2 domain containing genes. Panel A shows the RNA-Seq counts per million for AP2 and SPL genes. The top graph contains AP2 genes that are not regulated by miR172 while the bottom graph shows AP2 genes that are under regulation by the microRNA. The qPCR validation of some of the genes in panel A are shown in panel B along with the quantification of miR156 and miR172. Matching bar colors are used in both panels to indicate the identical genes.

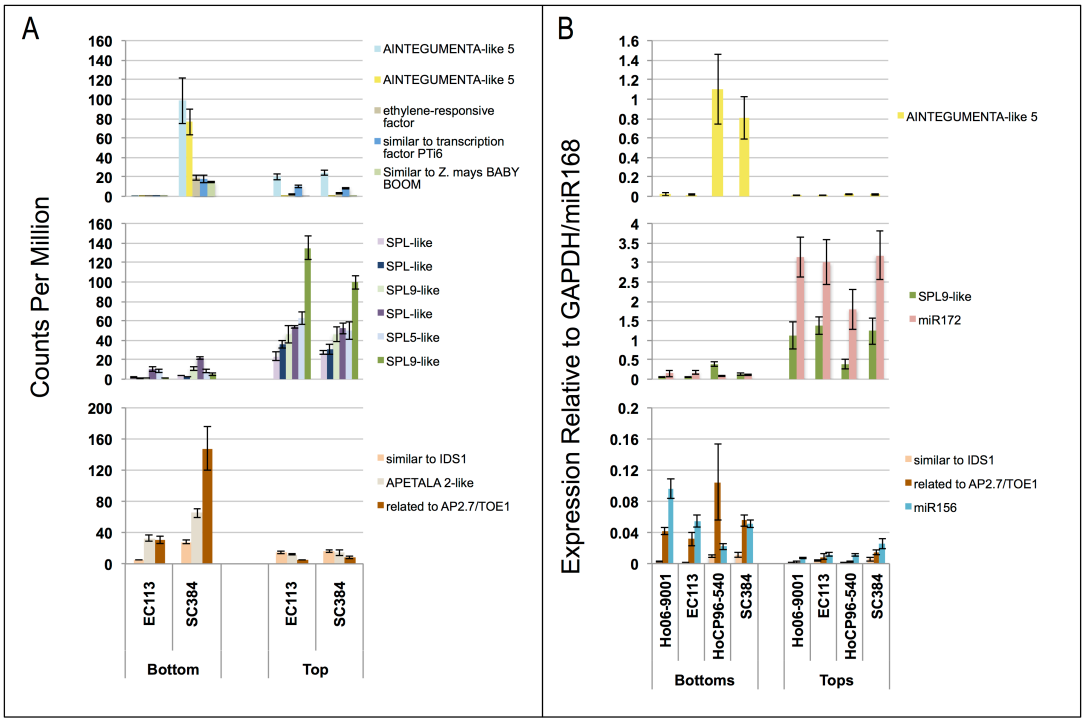


Figure 2.12: Expansin and *XTH* Expression in Stems. Four *XTH* genes, shades of brown, and four expansin genes, shades of blue, were differentially expressed between SC384 tissues and EC113 tissues. SC384 bottom internode expressed all eight genes at higher levels than EC113 bottoms.

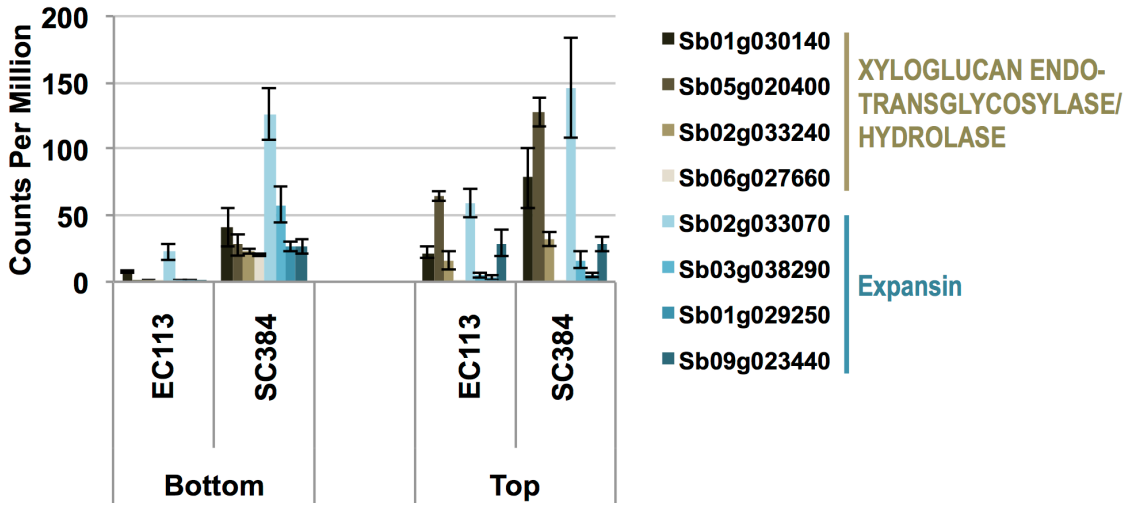
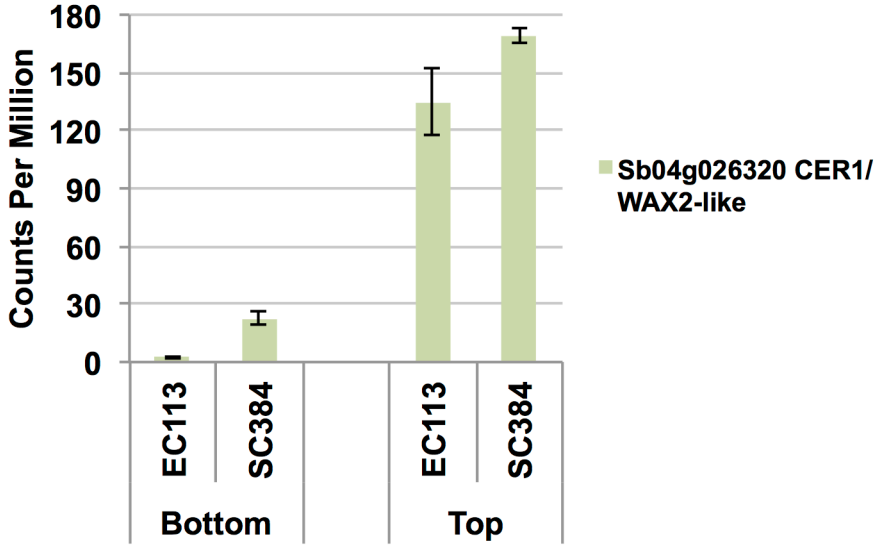


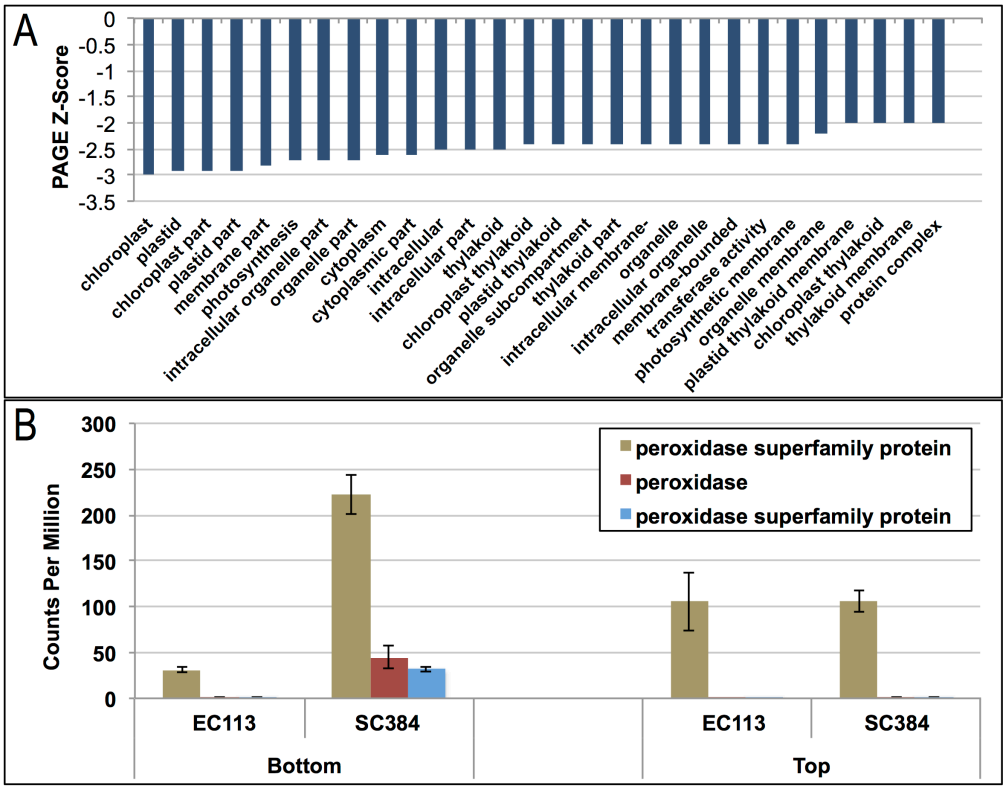
Figure 2.13: CER1/WAX2-like Expression. CER1/WAX2-like showed a nine-fold increase in expression between SC384 and EC113 bottoms. The tops had at least an 18-fold increase compared to their bottoms.



Of the 79 genes that are differentially expressed in both the comparison between the tops and bottoms of SC384 and EC113 (Figure 2.9, panel C), the majority of them

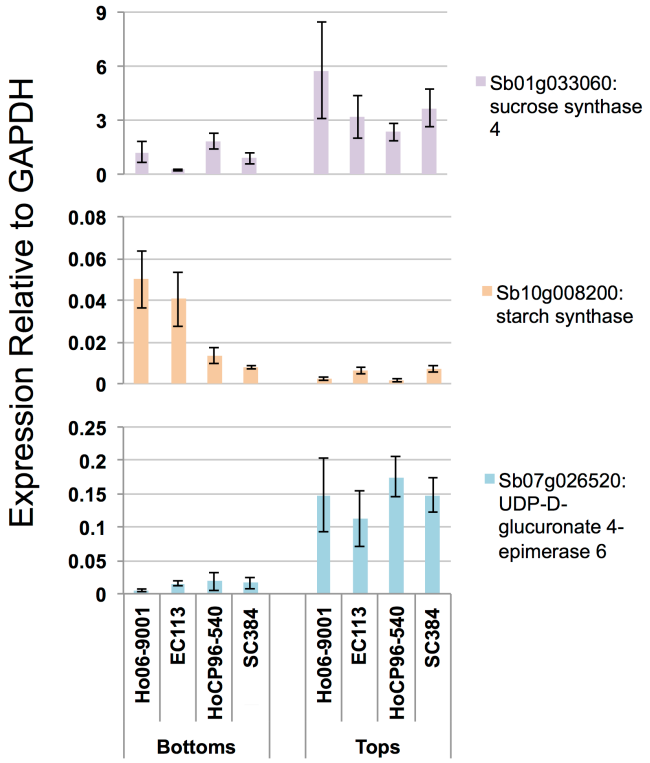
are related to photosynthesis. A Parametric Analysis of Gene Set Enrichment (PAGE) was performed using the Agrigo website (http://bioinfo.cau.edu.cn/agriGO/analysis.php?method=PAGE). A false discovery rate of 0.05 was used with minimum number of mapping entries set to 10. The results showed that all of the GO terms were photosynthesis related and they were all expressed at lower levels in SC384 bottoms compared to EC113 bottom internodes. The expression levels were not considered significant in the tops by the PAGE settings. SC384 bottoms also showed higher expression of peroxidase genes compared to both top tissues and EC113 bottoms (Figure 2.14).

Figure 2.14: GO Term Analysis and Peroxidase Expression. The top panel A shows the GO Term chart for the bottom tissue comparison that are also considered differentially expressed in the RNA-Seq analysis for the top tissues. Panel B shows that peroxidase expression is up in SC384 bottom tissues compared to all other tissues.



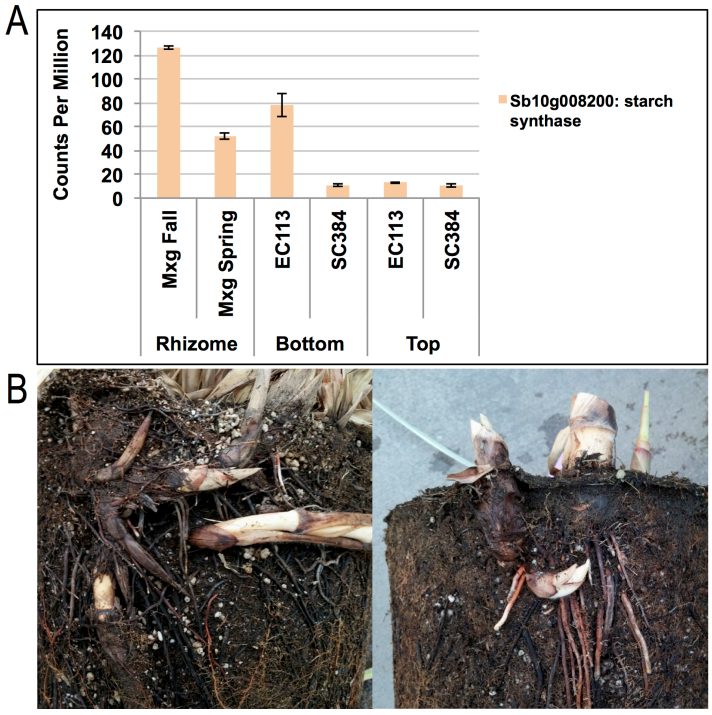
Three selected sucrose and starch metabolism genes, sucrose synthase 4, starch synthase, and UDP-D-glucoronate 4-epimerase 6, that were identified in the mRNA sequencing differential expression analysis as genes of potential interest were validated using qPCR. Sucrose synthase was more highly expressed in tops compared bottom internodes and higher in SC384 bottoms compared to the bottoms of EC113. However, the other high fiber variety tested, Ho 06-9001, showed no change in expression from the two high sucrose lines, SC384 and HoCP 96-540. Starch synthase showed elevated expressed in the bottoms of both high fiber *Saccharum* compared to the high sucrose bottoms and their tops. UDP-D-glucoronate 4-epimerase 6 was upregulated in the tops of all four lines compared to their bottoms (Figure 2.15).

Figure 2.15: Starch and Sucrose Metabolism Enzyme qPCR Analysis. The expression of three enzymes was quantified in two high fiber varieties, Ho 06-9001 and EC113, and two high sucrose varieties, HoCP 96-540 and SC384 relative to the expressed of GAPDH.



Since EC113 bottoms were shown to be expressing starch synthase at higher levels in both the RNAseq and qPCR result and EC113 was observed developing more tillers and small rhizomatous structures, a comparison to *Miscanthus* was made to test if a similar starch synthase was used by the closely related grass in its rhizome. Barling et al., 2013 compared fall and spring rhizomes of *Miscanthus* x *giganteus*. When we compare the counts per million output of Sb10g008200, starch synthase, from their fall and spring rhizomes to our top and bottom internode tissues, it shows that EC113 bottom internode expresses starch synthase at a level slightly higher than a spring rhizome but lower than a fall rhizome (2.16).

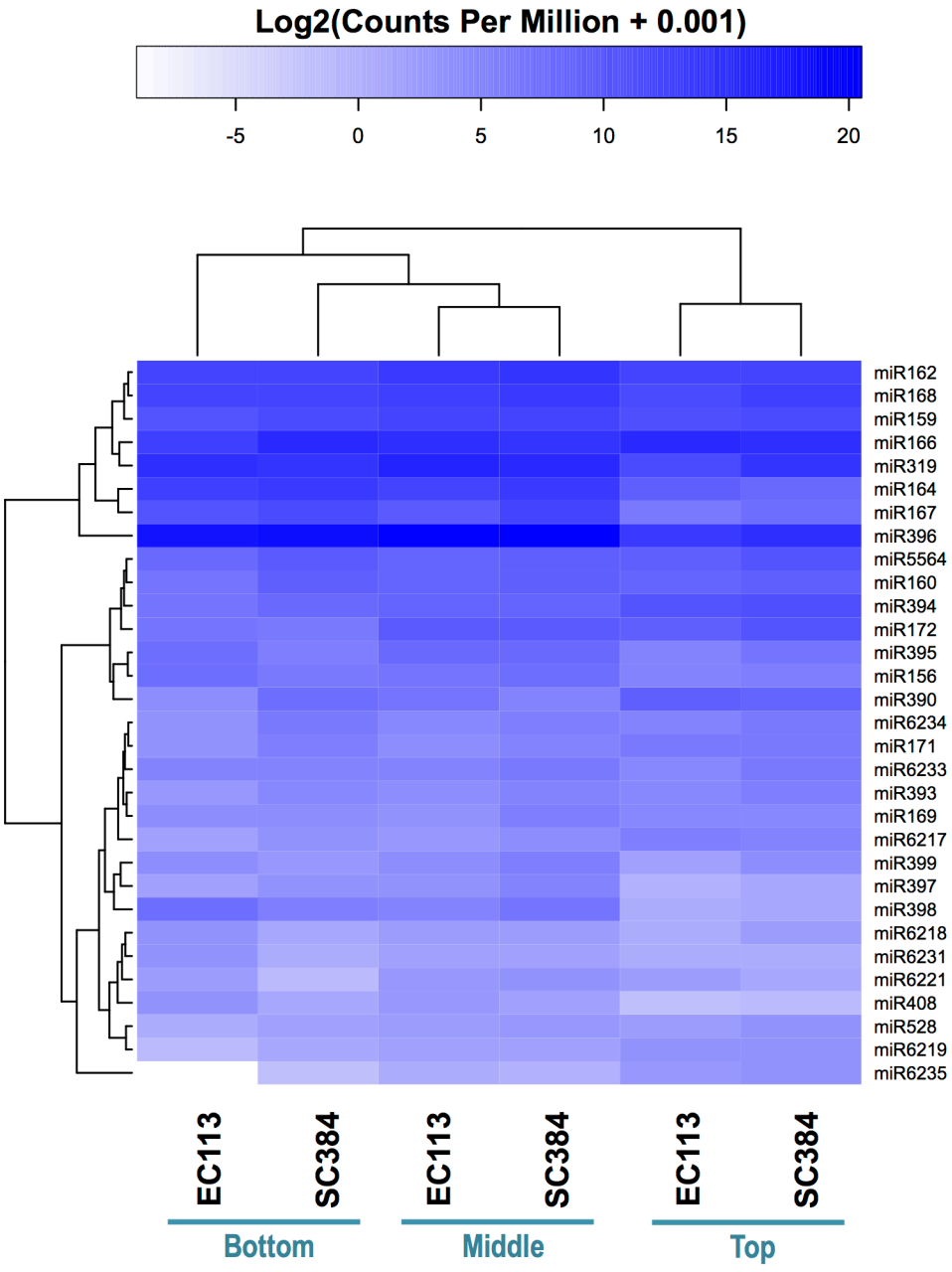
Figure 2.16: Starch Synthase Expression in *Saccharum* High Fiber and High Sucrose lines versus *Miscanthus* x *giganteous* spring and fall rhizome. Panel A shows RNAseq data from Barling et al. 2013 looking at spring and fall rhizomes was compared to our *Saccharum* stem internodes looking at the expression of a starch synthase. Expression is high in rhizomes and the bottom internode of EC113. EC113 rhizomatous structures are shown in panel B.



Expression of miR156, miR172, SPL gene Sb07g027740, and two AP2 domain containing genes with miR172 binding domains: Sb01g003400 similar to IDS1 and Sb09g002080 related to AP2.7/TOE1; were validated using qPCR. The results showed that the AP2 genes along with miR156 were expressed higher in the bottoms of both high fiber and high sucrose types while SPL and miR172 were higher in all top tissues. Similar to the RNAseq findings, Sb10g026150, AINTEGUMENTA-like 5, was only expressed in bottom internodes that contain higher levels of sucrose (Figure 2.11).

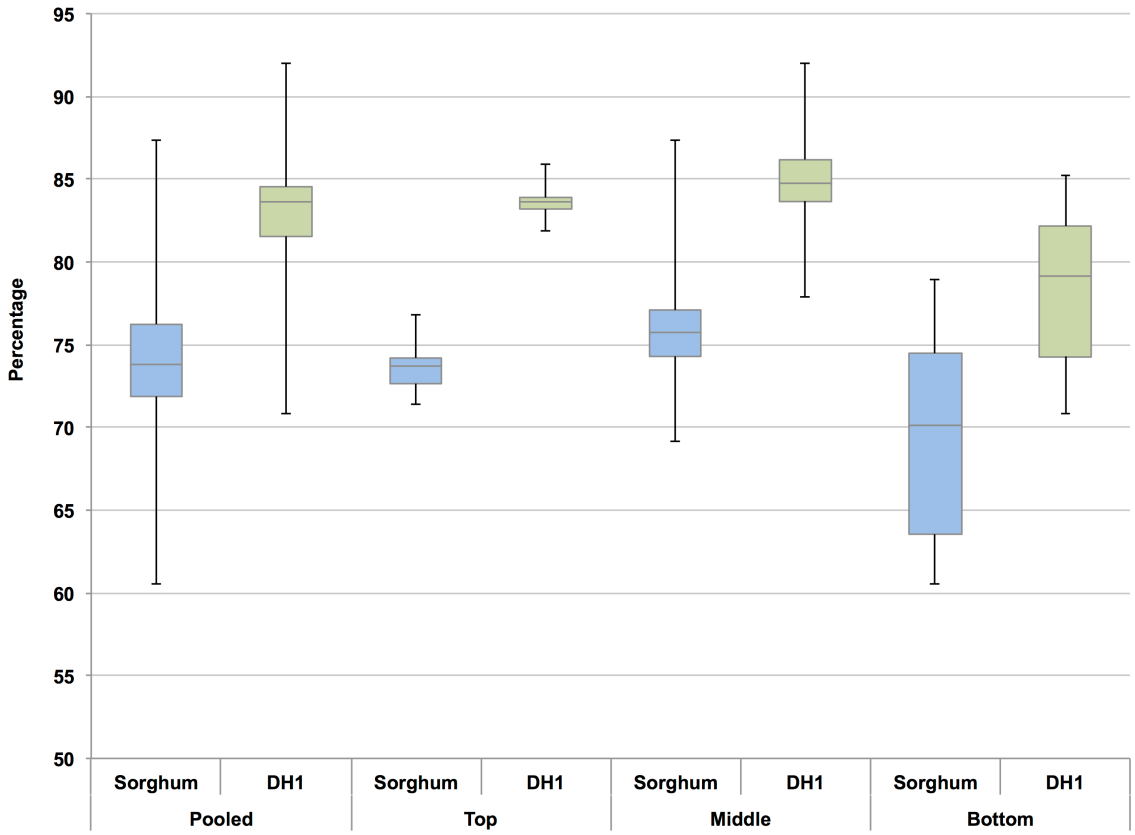
The heatmap produced from the miRNAseq counts shows that internode age is still the highest source of variance in the experiment with variety being second. However, the middle internode tissues seem to be more similar to the bottom internodes that the tops. The results for miR172 and miR156 are similar to the qPCR with miR172 being higher in the tops and miR156 being slightly higher in bottoms; however, middles seem to be highest for both mircorRNA (2.17).

Figure 2.17: miRNA Heatmap based on miRNAseq. The log base 2 values were taken from the miRNAseq count per million data. The clustering shows that internode age is still the major source of variance in the samples. Middle internodes are more similar to bottom internodes with SC384 being an intermediate between the middle internodes and EC113. The heatmap shows that miRNA172 and miR156 show a similar pattern of expression to that demonstrated by the qPCR.



Since *Miscanthus* and *Saccharum* are more closely related than *Saccharum* is to *Sorghum*, the *Saccharum* mRNA reads were also aligned to the DH1 version 3 assembly of *Miscanthus sinensis*. The mapping statistics were compared to the *Sorghum* aligned reads. About ten percent more reads mapped to DH1 overall than to sorghum when comparing statistics of all libraries. Tissue specific comparisons show that on average tops and middles had higher percentages of reads mapping and bottom tissue libraries not only had lower mapping rates but they were also more variable (Figure 2.18).

Figure 2.18: Sorghum versus DH1 Alignments. Overall mapping rates were pooled and compared for the sorghum and DH1 aligned libraries. When the libraries are separated by tissues but still pooling EC113 and SC384, it shows that tops and middles have higher mapping rates.



2.5 DISCUSSION

High fiber and high sucrose lines of *Saccharum* have similar within stem gene expression patterns as seen by the lower number of genes differentially expressed between SC384 and EC113 and that the majority of genes that are differentially expressed between the tops and bottom of the within variety comparisons overlap (Figure 2.9 panel B). However, the high sucrose, SC384, variety top and bottom tissues have fewer genes that differ indicating that they are more similar (Figure 2.9 panel B). The composition analysis shows that the SC384 tops and bottoms have similar levels of fiber but a different ratio of sucrose to other extractives (Figure 2.8). Sucrose synthase is expressed is higher in SC384 tops though (Figure 2.10, 2.15). Other studies have also shown higher sucrose synthase expression in young internodes of *Saccharum* stems and that the expression is negatively correlated with sucrose accumulation (Goldner et al., 1991; Schafer et al., 2004; Verma et al., 2011). The reaction of sucrose synthase works in both directions and it has been shown that the primary function in vivo is actually to degrade sucrose. The degradation products are then used for respiration and starch or cell wall biosynthesis (Lingle and Smith, 1991; Buczynski et al., 1993).

The higher expression levels of starch synthase in the bottoms of high fiber stems is interesting because there is not a large difference in starch accumulation in high fiber and high sucrose varieties. Starch would be a component of the hemi-cellulosic glucan (Figure 2.8). In *Miscanthus x giganteus (Mxg)*, stem sugar is converted to starch and sent to the rhizome in the fall when it flowers. During spring, that starch is remobilized to produce new tillers. EC113 is known to produce rhizomatous growth that contributes to profuse tillering (Gravois et al., 2012). When comparing expression of fall and spring rhizome tissues collected from *Mxg* (Barling et al., 2013) to the tops and bottoms of *Saccharum*, the same starch synthase gene was highly expressed in the rhizomes (Figure 2.16). This is interesting because perhaps the reason for the reduction in sucrose in high fiber bottoms is that they are converting their sucrose to starch and using it to generate rhizomatous tissue similar to their *Miscanthus*

relatives. It is also intriguing to note that EC113 and its energy cane siblings are known to be cold tolerant and very productive (Khan et al., 2013; Knoll et al., 2013, Friesen et al., 2014).

The increase in stem diameter in SC384 compared to EC113 could be due to the increased sucrose in the stems. Sugarcane has been shown to accumulate sucrose both inside and outside the cells (Welbaum & Meinzer, 1990) in stems. Using the older stem tissue as a sink could be limiting to the plant since stem size is determined before actual future sink necessity is known. SC384 showing higher expression of *AINTEGUMENTA*-like (*ANT*) genes in its bottom tissue might be a response to the increase in sucrose and the necessity for the plant to grow a larger sink (Figure 2.11). *ANT* has been shown to control plant organ cell number and organ size in *Arabidopsis* and tobacco shoots. Furthermore, *ANT* has been shown to affect development in mature organs. The researchers proposed that *ANT* maintains the meristemamic competence of cells (Mizukami & Fischer, 2000). *ANT* genes have also been linked to increased expression of expansin genes. Transgenic tobacco plants overexpressing ANT genes, PnANTL1 and PnANTL2, showed longer leaves (Kuluev et al., 2012). SC384 bottoms are also expressing higher levels of a gene similar to *BABY BOOM* (*BBM*). BBM is an AP2 domain containing gene that has also been linked to cell proliferation and embryogenesis in *Arabidopsis* (Boutilier et al., 2002). Expansins are believed to act in the relaxation of the cell wall. They may do this by breaking the bonds between cellulose microfibrils and matrix polysaccharides (McQueen-Mason and Cosgrove, 1995; Cosgrove et al., 2002) allowing for cell expansion. . Another cell wall targeting enzyme, *XTH*, can hydrolyse xyloglucans. Xylglucan is a major component of plant cell walls and transglycosylate residues into growing xyloglucan (Farrokhi et al., 2006). The LCP bottoms show increased expression of both XTH and expansin (Figure 2.12). There could be some correlation between the increase of *ANT* genes, expansin, and *XTH* in response to increased sugar. This response might act to create a larger sink for the plant and thus be the cause of the observed increase in cell size and stem diameter (Figure 2.8).

Another response to increased sucrose accumulation could be the down regulation of photosynthesis genes (Figure 2.14). Previous studies have shown that higher sugar in sugarcane has this effect on photosynthesis (McCormick et al., 2006, 2008). But something that is noticed when comparing the SC384 bottom to EC384 and SC384 top tissues is that there is an increase in peroxidase. It has been shown that when sugarcane is stressed that ascorbate peroxidase can improve the recovery of photosynthesis (Sales et al., 2013). It could be possible that the increase in peroxidase expression seen here is also in response to the downregulated photosynthesis and may be working to bring expression levels back up.

It is interesting to note that expression of a *CER1/WAX2*-like gene is expressed higher in SC384 bottoms compared to EC113. In *Arabidopsis*, *CER1* and *WAX2* have been shown to be involved in the formation of epicuticular wax (Aarts et al., 1995; Chen et al., 2003). SC384 is described as producing excessive amounts of wax while EC113 only produces moderate wax (Gravois et al., 2012). AP2 domain transcription factors have been shown to activate wax biosynthesis (Aharoni et al., 2004; Broun et al., 2004; Moose and Sisco, 1994). Overexpression of one of these AP2 genes, *GLOSSY15* (*gl15*), led to an increase in stem sugars (Chapter 1). SC384 bottoms show an increase in expression of AP2 domain containing genes. Three of them are similar to *gl15*. The accumulation of wax and higher similarity of SC384 bottoms to its top than EC113 bottom could be could be an indication that the bottoms of SC384 are behaving like younger tissue. This could be a response of sugarcane to the increase in AP2 genes, similar to *gl15*, which are known to promote the juvenile phase.

The higher alignment rates (Figure 2.15) of *Saccharum* reads to the *Miscanthus* DH1 assembly compared to the sorghum genome alignments is not surprising considering *Miscanthus* and *Saccharum* are more closely related (Lawrence & Walbot, 2007; Kim et al., 2014). *Miscanthus* has a whole genome duplication (Swaminathan et al., 2012) and *Saccharum* is also a polyploid. The sorghum genome is based on a diploid. Aligning to DH1 might allow for a separation of paralogs that sorghum cannot

detect. A completed DH1 genome would be a powerful tool that could help unravel the more complex *Saccharum* genome.

Further work is needed to understand the entire mechanism for the differences in stem carbon partitioning in *Saccharum* that causes high fiber and high sucrose types. Transgenics could be created which knockout AP2 genes to see if that creates a plant that is more similar to the low sucrose *S. spontaneum*. Also, the role of *AINTEGUMENTA* in sugarcane sink expansion needs to be further explored.

2.6 REFERENCES

Aarts, M.G., Keijzer, C.J., Stiekema, W.J., & Pereira, A. (1995) Molecular characterization of the CER1 gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* **7**, 2115–2127.

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., van Arkel, G., & Pereira, A. (2004) The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. *The Plant Cell* **16**, 2463–2480.

Anders, S., Pyl, P.T., & Huber, W. (2014) HTSeq — A Python framework to work with high-throughput sequencing data. *bioRxiv* preprint, doi: 10.1101/002824.

Barling, A., Swaminathan, K., Mitros, T., James, B.T., Morris, J., Ngamboma, O., Hall, M.C., Kirkpatrick, J., Alabady, M., Spence, A.K., Hudson, M.E., Rokhsar, D.S., & Moose, S.P. (2013) A detailed gene expression study of the *Miscanthus* genus reveals changes in the transciptome associated with the rejuvenation of spring rhizome. *BMC Genomics* **12**, doi:10.1186/1471-2164-14-864.

Broun, P., Poindexter, P., Osborne, E., Jiang, CZ., & Riechmann, J.L. (2004) WIN1, a transcriptional activator of epidermal wax accumulation in Arabidopsis. *PNAS* **101**, 4706-4711. doi:10.1073/pnas.0305574101.

Bolger, A.M., Lohse, M., & Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* btu170.

Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C.-M., van Lammeren, A.A.M., Miki, B.L.A., Custers, J.B.M., & van Lookeren Campagne, M.M. (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **14**, 1737–1749.

Buczynski, S.R., Thom, M., Chourey, P., & Maretzki, A. (1993) Tissue distribution and characterization of sucrose synthase isozymes in sugarcane. *J. Plant Physiol*. **142**, 641–646.

Cardoso-Silva, C.B., Costa, E.A., Mancini, M.C., Balsalobre, T.W., Canesin, L.E., Pinto, L.R., Carneiro, M.S., Garcia, A.A., de Souza, A.P., & Vicentini, R. (2014) De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS One* **9**, e88462.

Carpita, N.C. & Gibeaut, D.M. (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J*. **3**, 1-30.

Chen, X., Goodwin, S.M., Boroff, V.L., Liu, X., & Jenks, M.A. (2003) Cloning and characterization of the WAX2 gene of Arabidopsis involved in cuticle membrane and wax production. *The Plant Cell* **15**, 1170–1185.

Correns, C. (1901) Bastarde zwischen maisrassen, mit besonder Berucksichtung der Xenien. *Bibliotheca Bot*. **53**, 1–161.

Cosgrove, D.J., Li, L.C., & Cho, H.T., Hoffmann-Benning, S., Moore, R.C., Blecker, D. (2002) The growing world of expansins. *Plant and Cell Physiology* **43**, 1436–1444.

de Setta, N., Monteiro-Vitorello, C.B., Metcalfe, C.J., Cruz, G.M., Del Bem, L.E., Vicentini, R., Nogueira, F.T., Campos, R.A., Nunes, S.L., Turrini, P.C., Vieira, A.P., Ochoa Cruz, E.A., Corrêa, T.C., Hotta, C.T., de Mello Varani, A., Vautrin, S., da Trindade, A.S., de Mendonça Vilela, M., Lembke, C.G., Sato, P.M., de Andrade, R.F., Nishiyama Jr, M.Y., Cardoso-Silva, C.B., Scortecci, K.C., Garcia, A.A., Carneiro, M.S., Kim, C., Paterson, A.H., Bergès, H., D'Hont, A., de Souza, A.P., Souza, G.M., Vincentz, M., Kitajima, J.P., & Van Sluys, M.A.. (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* **15**, doi: 10.1186/1471-2164-15-540.

FAOSTAT. “Download Data” Retrieved May 25, 2013, from http://faostat3.fao.org/home/index.html#DOWNLOAD.

Farrokhi, N., Burton, R.A., Brownfield, L., Hrmova, M., Wilson, S.M., Bacic, A., & Fincher, G.B. (2006) Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. *Plant Biotechnology Journal* **4**, 145-167.

Friesen, P.C., Peixoto, M.M., Busch, F.A., Johnson, D.C., & Sage, R.F. (2014) Chilling and frost tolerance in Miscanthus and Saccharum genotypes bred for cool temperate climates. *J. Exp. Bot*., eru105.

Goldner, W., Thom, M., & Maretzki, A. (1991) Sucrose metabolism in sugarcane cell suspension cultures. *Plant Sci*. **73**, 143–147.

Grativol, C., Regulski, M., Bertalan, M., McCombie, W.R., da Silva, F.R., Zerlotini Neto, A., Vicentini, R., Farinelli, L., Hemerly, A.S., Martienssen, R.A., & Ferreira, P.C. (2014) Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus Saccharum. *Plant J.* **79**, 162-72.

Gravois, K., Fontenot, D., Pontif, M., Kimbeng, C., Waguespack, H., Dufrene, E., & Duet, M. (2012) Louisiana Sugarcane Variety Identification Guide. (Pub.3056) Baton Rouge, Louisiana: LSU AgCenter Research & Extension.

Jacobsen, K.R., Fisher, D.G., Maretzki, A., & Moore, P.H. (1992) Developmental Changes in the Anatomy of the Sugarcane Stem in Relation to Phloem Unloading and Sucrose Storage. *Botanica Acta*. **105**, 70–80.

Khan NA, Bedre R, Parco A, Bernaola L, Hale A, Kimbeng C, Pontif M, Baisakh N. (2013) Identification of cold-responsive genes in energycane for their use in genetic diversity analysis and future functional marker development. Plant Science 211, 122–131.

Kim, C., Wang, X., Lee, T.H., Jakob, K., Lee, G.J., & Patterson, A. (2014) Comparative Analysis of *Miscanthus* and *Saccharum* Reveals a Shared Whole-Genome Duplication but Different Evolutionary Fates. *Plant Cell* **26**, 2420-2439.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36.

Knoll, J.E., Anderson, W.F., Richard Jr, E.P., Doran-Peterson, J., Baldwin, B., Hale, A.L., & Viator, R.P. (2013) Harvest date effects on biomass quality and ethanol yield of new energycane (Saccharum hyb.) genotypes in the Southeast USA. *Biomass and Bioenergy* **56**, 147–156.

Kuluev, B.R., Knyazev, A.V., Iljassowa, A.A., & Chemeris, A.V. (2012) Ectopic expression of the PnANTL1 and PnANTL2 black poplar genes in transgenic tobacco plants. *Russ J Genet* **48**, 993–1000.

Langdon, L.M. (1920) Sectioning Hard Woody Tissues. *Botanical Gazette* **70**, 82-84.

Lawrence, C.J., and Walbot, V. (2007). Translational genomics for bioenergy production from fuelstock grasses: maize as the model species. Plant Cell 19: 2091–2094.

Lingle, S. E. & Smith, R. C. (1991) Sucrose metabolism related to growth and ripening in sugarcane internodes. *Crop Sci*. 31, 172–177.

McCormick, A.J., Cramer, M.D., & Watt, D.A. (2006) Sink strength regulates photosynthesis in sugarcane. *New Phytologist* **171**, 759-770.

McCormick, A.J., Cramer, M.D., & Watt, D.A. (2008) Regulation of photosynthesis by sugars in sugarcane leaves. *Journal of Plant Physiology* **165,** 1817-1829.

McQueen-Mason, S.J. & Cosgrove, D.J. (1995) Expansin Mode of Action on Cell Walls (Analysis of Wall Hydrolysis, Stress Relaxation, and Binding). *Plant Phy*. **107**, 87-100.

Mizukami, Y. & Fischer, R.L. (2000) Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proc Natl Acad Sci USA*. **97**, 942–947.

Moose, S.P. & Sisco, P.H. (1994) *Glossy15* Controls the Epidermal Juvenile-to-Adult Phase Transition in Maize. *Plant Cell* **6**, 1343-1355.

Pan, D. & Nelson, O.E. (1984) A debranching enzyme deficiency in endosperms of the sugary-1 mutants of maize. *Plant Physiol*. **74**, 324–328.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., ... & Peterson, D. G. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**, 551-556.

Ploner, A. (2014) *Heatplus: Heatmaps with row and/or column covariates and colored clusters*. R package version 2.12.0.

Reiter, W.D. (1998) The molecular analysis of cell wall componets. *Trends Plant Sci*. **3**, 27-32.

Richard Jr, E.P., Tew, T.L., Cobill, R.M., & Hale, A.L. (2008) Sugar/Energy Canes as Feedstocks for the Biofuels Industry. Proc. Short Rotational Crops International Conference, August 19-21, 2008, Bloomington, Minnesota. p. 47.

Sales, C.R., Ribeiro, R.V., Silveira, J.A., Machado, E.C., Martins, M.O., & Lagôa, A.M. (2013) Superoxide dismutase and ascorbate peroxidase improve the recovery of photosynthesis in sugarcane plants subjected to water deficit and low substrate temperature. *Plant Physiol Biochem*. **73**, 326-236.

Schafer, E.W., Rohwer, J.M., & Botha, F.C. (2004) Protein level expression and localization of sucrose synthase in sugarcane culm. *Physiol. Plant*. **121**, 187–195.

Schnable, P.S., Ware. D., Fulton, R.S., Stein, J.C., Wei, F, et al. (2009) The B73 maize genome: complexity, diversity and dynamics. *Science* **326**: 1112–1115.

Schwarz, K-U, Kjeldsen, J.B., Munzer, W., & Junge, R. Low cost establishment and winter survival of Miscanthus x giganteus. In: Kopetz H, Weber T, Palz W, Chartier P. Ferrero GL, editors. Biomass for energy and the environment: Proceedings of the 10th European Bioenergy Conference, Würzburg, Germany, 8–11 June 1998. Rimpar, Germany: C.A.R.M.E.N., 1998. p. 947–50.

Scurlock, J.M.O. (1998) Miscanthus: a review of European experience with a novel energy crop. ORNL/TM-13732. Oak Ridge National Laboratory, Oak Ridge, Tennessee. 26 pp.

Swaminathan, K., Chae, W.B., Mitros, T., Varala, K., Xie, L., Barling, A., Glowacha, K., Hall, M., Jezowski, S., Ming, R., Hudson, M., Juvik, J.A., Rokhsar, D.S., & Moose, S.P. (2012) A framework genetic map for Miscanthus sinensis from RNAseq-based markers shows recent tetraploidy. *BMC Genomics*. **13**, 142.

Verma, A.K., Upadhyay, S.K., Verma, P.C., Solomon, S., & Singh, S.B. (2011) Functional analysis of sucrose phosphate synthase (SPS) and sucrose synthase (SS) in sugarcane (Saccharum) cultivars. Plant Biol. 13 325–332.of solutes and water in developing sugarcane stalk tissue. *Plant Physiol*. **93**, 1147–1153.

White, W., Vincent, M., Moose, S.P., & Below, F. (2012) The sugar, biomass and biofuel potential of temperate by tropical maize hybrids. *GCB Bioenergy* **4**, 496-508.